

# Confusing tale of two research budgets

**The fashion for strategic research that stamps this year's US budget echoes Britain's passion for wealth-creation — and may similarly begin to fade as the difficulties become apparent.**

ALL research is strategic now, or at least must be regarded in that light. That is the message from Washington this week (see page 497), where the Clinton administration's second budget was published on 7 February. Gone are the days when the United States supported science for the prestige that would accumulate as a consequence. The decision of the US Congress last year not to continue building the Superconducting Super Collider (SSC) was a sufficient proof of that. What has changed? In the old days, the Congress would have been alarmed that the abandonment of SSC would have handed 'leadership' in particle physics to the Soviet Union. But the Cold War is now ended and, with it, the sense that competition even in such abstract fields may be a productive extravagance. The consequences for the financing of research are widespread and profound (see *Nature* 367, 5; 6 January 1994). Who would expect the United States to be unaffected?

The meeting organized in Washington last week by the Office of Science and Technology Policy showed which way the wind is blowing, but also pointed to potential divisions between the research community and the US government. The Congress and the administration are evidently at one in their conviction that the urgent need is for domestic economic growth ('jobs'), to which end successful competition on international markets in technically novel goods and services has become indispensable. Moreover, they ask that science should be harnessed to that goal, whose attainment is the stuff of which successful re-election campaigns are made. Powerful Senator Barbara A. Mikulski put the point plainly last week (see page 497) with her repetition of her theme for the past several months that research must create "new technologies which must lead to jobs, particularly in manufacturing".

Last week's meeting may therefore have been a rehearsal for the arguments there will be in Congress over the research lines in the federal budget during the forthcoming budget cycle. Indeed, the administration's request for funds shows some signs of trimming to the current mood. The budget of the National Aeronautics and Space Administration (NASA) will actually fall under the administration's plans, for example, which is a realistic acknowledgement that even the strongest supporters of the involvement of Russia in the new-style space station cannot pretend that the outcome will benefit competitiveness and increase domestic employment.

That is how the meaning of "strategic" must be understood. It is not an unfamiliar cry, witness last year's debate

in Britain about the harnessing of research to "wealth creation". Nor, of course, is the objective in any way disreputable; what scientist would not be cheered to know that his or her research won practical benefits for the wider world as well as a modicum of understanding? The difficulties are those of telling in advance which particular pieces of research will lead to 'new technologies' and then to 'jobs'.

The recent past is littered with examples of adventurous goal-directed programmes of research and development which have failed for intrinsic reasons or which, alternatively, have been technically successful, but unusable for economic or other reasons. Of course, it is usually possible to identify technical fields in which bets are safer. More research on earthquake mechanisms and on the seismic stability of buildings would be widely welcomed, especially in California (and might even create jobs in the construction industry), but this is not a field in which the United States will be competing fiercely with Japan and its industrial successors decades from now. Nor will the US effort in understanding climate change, valuable though that is, contribute much to these economic goals.

A further complication is that the most productive innovations usually destroy jobs, and create them only indirectly, by liberating human resources for other tasks (and only then if the competitive advantages they create work through into the general prosperity). Between them, the steam engine, the railway locomotive and the automobile put an end to the ancient occupation of the farrier, just as the arrival of the personal computer is likely to see off the trade of stenogra-

## N&V competition

As part of its 125th anniversary celebrations, *Nature* plans to offer two prizes to contributors to its News and Views section. One prize of £500 will be awarded to the author of the most interesting and literate contribution solicited in the usual way, the other (for the same amount) for the most interesting and literate unsolicited contribution. All articles published between 4 November 1993 and 3 November 1994, except those by members of the staff of *Nature*, will be considered. The decision by the News and Views Editor, Tim Lincoln, will be final, and will be announced on 3 November. **Those wishing to offer unsolicited articles are urged to seek advice from Tim Lincoln in advance.** □



pher. There is no surprise in that. Has not Daniel Bell's "post-industrial society" been with us for a quarter of a century already? Recent experience suggests a stark conflict between the laudable goals of technological competitiveness and reasonably full employment. To pretend that it can be otherwise may be to believe that water can spontaneously run uphill.

To be fair, there are signs that the Congress understands the difficulties. Mikulski acknowledged last week that every research grant need not yield "six patents and four commercial licensing agreements". The research community's best hope of winning friends in Congress without abandoning its cry (repeated last week) for continued support for familiar research mechanisms and institutions is to proclaim (preferably with supporting documentation) that the research enterprise is important to the US economy as the only means of creating the human cleverness from which radical innovations spring. Vice-President Al Gore's advocacy of the electronic superhighway as a means of making the United States as a whole intellectually more productive is a more populist route to the same objective.

Those are some of the reasons why the fashionable US affection for strategic research may prove short-lived. In Britain, much the same seems to be happening. Having pinned its reorganization of research on the doctrine of science for wealth-creation, the government appears now to be more conscious of the problems it has undertaken to solve. Indeed, the prime minister, John Major, seemed to be suggesting in a speech last week that the British part of the research enterprise deserves respect of the kind accorded to other social institutions at the heart of his "back-to-basics" rhetoric. After more than a decade of needless damage-doing, that would be only prudent.

The chief consequences for British research in the immediate future (apart from the reorganization of the research councils, see page 501) are likely to be modest spending outside the standard grant-making rules of some £15 million a year, or 5 per cent of the total spending on civil science. It may be possible to guess in April whether the 'foresight' exercise, which the government hopes will allow researchers' efforts to be more productively directed, will indeed have that effect, but changes in the pattern of research activity will become apparent only over several years. Meanwhile, the government is in a quandary over the future management of Britain's largest civil laboratory, the Rutherford Appleton Laboratory, originally set up to provide central research services for academic research groups. The courageous, but also wise, course would be to invite a consortium of research universities to run the place.

On the grander questions, on both sides of the Atlantic, it seems likely that the first flush of enthusiasm for turning research into prosperity will be abated by the reality of the difficulties of doing so. When governments discover in the source of seeking radical reorganization that the best they can do with their parts of the research enterprise is to cherish them, the lessons are likely to be remembered. If the outcome in the research community is a more vivid awareness of how much the world at large looks to research for its improvement, so much the better. □

## Ecumenical curricula

**Religious instruction cannot be at once equitable among religious beliefs and also evangelical.**

STATESMEN the world over know that illogical policies cast a shadow over policy-making for years ahead. There is a splendid illustration of this principle in the contortions the British government is having to undergo to adapt the school curriculum to modern needs, and those of religious instruction in particular. Even flamboyant Henry VIII, the sixteenth century monarch, would be surprised that his determination to match the popes by having a church of his own, of which he would be the spiritual head, would remain an embarrassment after more than 400 years. But that is what is happening.

At the end of last month, Baroness Blatch, the government's sensible education minister in the House of Lords, did her best to square the circle bequeathed by Henry VIII in defending what are called "guidelines" for religious instruction in British schools. Word has gone out (beneath official letterheads) that the bulk of the time allotted to religious instruction should be devoted to the Christian religion, but that all children should be taught about two other religious faiths by the time they are eleven, about a further two by fourteen and about "all five major non-Christian" religions by the end of compulsory schooling at sixteen. The decision seems to have maddened all concerned. People who are Moslems, for example, are asking for equal time with Christianity, but Christians who look to the schools for support for their faith fear the dilution of the message by the prescribed parade of ecumenical principles.

Much has changed since the sixteenth century, when Henry set up the Anglican Church as Britain's spiritual guardian (continuing the intolerant marginalization of those of Jewish faith in the process). Now there is a large company of Moslems, Sikhs and Hindus as well, most with voting rights. British Buddhists are relatively inconspicuous, while the company of the irreligious is uncounted. That is why it has become impossible to make the schools simple defenders of the Christian faith, but the Christians are certainly correct to fear that the advertisement the schools will now provide for religions other than their own, even for a small part of the time available, will blunt the sense of conviction in revealed truth that seems inseparable from religious belief.

The logical resolution of this dilemma would be to secularize the schools (which would incidentally help in sectarian Northern Ireland), but that may be too much to ask when the British government promises a "return to basics". Moreover, Christian sects were among the first to recognize the British need for public education in the nineteenth century; their foresight is still repaid with influence. Blatch's ecumenical curriculum, by compelling even believing teachers to deal with religion as if it were somewhere between history and sociology, may be the next best thing. □



## Senators warn of changing climate for research

**Washington.** The chairman of the Senate subcommittee responsible for approving a large proportion of the annual US science budget suggested last week that, as industry is the main "customer" for much university-based research, industrial representatives should play a greater role in deciding how federal research funds are distributed.

The suggestion came from Senator Jay Rockefeller (Democrat, West Virginia) who chairs the subcommittee on science, technology and space. He was speaking at a two-day meeting organized by the Office of Science and Technology Policy (OSTP), widely seen as part of the preparation for an anticipated presidential white paper on how the United States can maintain its world leadership in basic science.

Rockefeller suggested that industry might increase its influence over the science budget through, for example, greater representation on groups such as the National Science Board, the body that oversees the National Science Foundation (NSF). He also warned that if the US Senate votes on 22 February in favour of an amendment to the Constitution requiring the United States to balance its budget — as the House of Representatives has already done — the move could mean the end for new science projects.

The OSTP meeting provided an opportunity for more than a hundred researchers from academic institutions and industry to come together with the heads of most of the federal science agencies to talk about the importance of unfettered, curiosity-driven research.

Unsurprisingly, the scientists' main message was that the scientific enterprise needs more money. But the members of Congress who attended the meeting brought a different message, namely that future funding for science will depend on its success in creating jobs. This, for example, was the main theme of an address by Senator Barbara A. Mikulski (Democrat, Maryland) whose subcommittee controls the budgets of a number of science-based agencies, including the National Aeronautics and Space Administration and the NSF.

**Barbara Culliton**



**Mikulski**

## Clinton budget proposes science funding freeze

**Washington.** Small increases in science funding for the 1995 fiscal year, which starts this October, were proposed by President Bill Clinton when he presented his second budget to Congress on Monday (7 February). But the increases will be sufficient only to keep up with the anticipated rate of inflation.

The National Science Foundation (NSF) fares well in the submission, and the National Institutes of Health (NIH) less so, while both the National Aeronautics and Space Administration (NASA) and the Department of Energy face substantial cuts in parts of their research programmes.

The biggest increases will go to research related to new civilian technology initiatives, such as the planned 'information superhighway'. As a result, funding for applied research will rise slightly more than that for basic research.

The president proposes a 3.7 per cent increase in the federal research and development budget in 1995 compared to 1994 (excluding new research facilities in each case) to just over \$71 billion (see table). But after deep cuts in funding for new facilities, the increase would be only 2.7 per cent, probably close to the rate of inflation.

The proposed increases come as part of the first budget in which the total amount of 'discretionary spending' — that which the federal government is not obliged to make under the law — has been frozen in dollar terms by congressional decree. John Gibbons, the president's science adviser, says the increases therefore show that Clinton is "willing to cut other programmes to pay for research and development".

But Congress has yet to make its mark on the budget. And as proposed cuts in such politically sensitive areas as public housing work their way through the House of Representatives and the Senate, the modest increase for science is likely to be driven substantially below the expected rate of inflation.

The budget proposes no shift in the balance between civil and military research, with the latter still consuming 53 per cent of the total. Officials say spending has moved more quickly than planned in favour of civil research during Clinton's first year, and that the administra-

tion is still aiming for parity between the two by 1998. But liberals will be dismayed that Clinton plans to increase defence research at the same 4 per cent rate as health research.

Technology transfer programmes organized by the National Institute of Standards and Technology — part of the Commerce Department — get the sharpest boost of all, jumping from \$490 million this year to \$874 million in 1995, and programmes in various agencies related to the information superhighway being promoted by Vice-President Al Gore will rise from \$964 million to \$1,272 million.

Co-operative Research and Development Agreements (CRADAs) between government laboratories and private industry are also singled out for expansion.

**Colin Macilwain**

### NSF still keeps favoured agency status

Monday's budget proposals have confirmed the favoured status that the NSF enjoys under the Clinton administration. In particular, the NSF is the only federal agency whose entire workload is identified by science adviser Jack Gibbons as a "priority area".

The administration has requested a 6 per cent increase in NSF's overall funding, which will reach \$3.2 billion in the fiscal year beginning on 1 October, with the research component growing by more than 8 per cent to reach \$2.35 billion.

"It's a very good budget in a very tough budget year," says Neal Lane, the director of NSF, adding that the entire budget ➤

#### Proposed increases by agency

| Research and development (R&D):<br>(Budget authority)*<br>\$ millions | 1994 enacted       | 1995 proposed      | % Change 1994 to 1995 |
|-----------------------------------------------------------------------|--------------------|--------------------|-----------------------|
| Defense                                                               | 35,538             | 36,971             | 4%                    |
| Health and Human Services<br>(NIH)                                    | 11,033<br>(10,486) | 11,484<br>(10,994) | 4%<br>(5%)            |
| Commerce                                                              | 919                | 1,204              | 31%                   |
| NSF                                                                   | 2,026              | 2,220              | 10%                   |
| NASA                                                                  | 8,493              | 8,597              | 1%                    |
| Transportation                                                        | 617                | 692                | 12%                   |
| EPA                                                                   | 536                | 582                | 9%                    |
| Agriculture                                                           | 1,393              | 1,394              | 0%                    |
| Energy                                                                | 6,054              | 6,052              | -0%                   |
| Other                                                                 | 1,876              | 1,833              | -2%                   |
| <b>Total R&amp;D</b>                                                  | <b>68,484</b>      | <b>71,029</b>      | <b>4%</b>             |

( Figures exclude fundings for research facilities)



submission was consistent with a recent request from a congressional committee that the agency should place more emphasis on so-called 'strategic' research.

No absolute cut is proposed for the funding of any scientific discipline. But some 'pure' disciplines are promised only moderate expansion — astronomy, for example, would receive a 5 per cent spending increase — while those closer to policy goals, such as social science and some branches of computer science, would get increases of 20 per cent or more.

**C. M.**

## NIH: better than last year

The administration's budget request for the NIH for the fiscal year 1995 is nothing to shout about. But in the face of tight controls on federal spending overall, the situation could have been worse. President Clinton is requesting \$517 million more in the year beginning on 1 October for NIH than in the current year, a modest increase of 4.7 per cent that would bring its budget to \$11.5 billion.

Apart from a few targeted areas — research on women's health, and in particular breast cancer, HIV/AIDS and tuberculosis, support for high-performance computing and the minority health initiative, which are all healthy increases — funding for other research at NIH's 22 institutes and centres will remain level in real terms, the extra money amounting to little more than a cost-of-living increase.

Strong emphasis has been placed on research that is focused on the prevention of disease. Interestingly, however, the administration has frozen almost the entire budget for the Centers for Disease Control and Prevention, based in Atlanta, Georgia, which is the leading public health agency in the United States.

In addition to the \$11.5 billion requested for NIH, the administration has recommended that it should receive an additional \$400 million for research into prevention as part of the president's request for fiscal year 1995 for health-care reform.

Considering that this is a stringent budget overall, the president's request for NIH is better this year than last, says Harold Varmus, the director of NIH. "As long as we [NIH] are keeping ahead of the biomedical research and development price index, we're doing pretty well."

**Diane Gershon**

## Energy: cuts for nuclear

With a request of \$18.5 billion for the 1995 fiscal year, the Department of Energy (DoE)'s budget is proposed by the Clinton administration to fall 3 per cent for the second year in a row.

The largest cuts proposed by the

administration are in defence-related programmes (down 13 per cent, to \$5.6 billion) and nuclear energy research (down 25 per cent, to \$248 million). DoE plans to use some of these savings to increase spending on renewable energy and energy efficiency programmes, which would jump from a combined \$1 billion this year to \$1.4 billion in 1995.

The \$6.3-billion request for environmental cleanup at former defence laboratories is roughly the same as this year's spending level, and remains the largest single item of expenditure in the department's budget.

The Office of Energy Research is requesting \$2.8 billion, down 15 per cent from the amount finally allocated to it in 1994 (largely as a result of the phasing-out of the Superconducting Super



Space station funding remains at this year's level

Collider). More than a quarter of that amount goes towards DoE's programme of wide-ranging research in basic energy sciences, which is asking for slightly less than it received this year.

The 1995 request includes money to begin construction of the Tokamak Physics Experiment facility for fusion research at Princeton University, as well as the Advanced Neutron Source at Oak Ridge National Laboratory in Tennessee. It also continues funding for a variety of physics projects, such as the Advanced Photon Source, the B-Factory at Stanford University and the Relativistic Heavy Ion Collider. Funding for DOE laboratories will remain about level, although some will absorb significant cuts.

In the area of nuclear energy, the department plans to terminate research on two types of reactors deemed to have no near-term commercial return — the Advanced Liquid Metal Reactor and Modular High-Temperature Gas-Cooled Reactor. The Actinide Recycle Program also will be terminated. But research and development work on light water reactors will continue.

**Tony Reichhardt**

## EPA: bigger budget and more jobs

After taking criticism from environmentalists for cutting last year's budget for the Environmental Protection Agency (EPA), the Clinton administration is proposing to reverse this with an 8 per cent increase to \$7.2 billion — the largest request in the agency's history.

Funding for the EPA's Office of Research and Development would rise 7 per cent to \$570 million, with an increased emphasis on ecosystems research. The agency also plans to convert 900 contractor positions to government employee status in 1995; in a move to boost its in-house scientific and technical expertise in-house.

**T. R.**

## NASA: science up

NASA will see its first real budget reduction since the early 1970s, with a request of \$14.3 billion. This is \$965 million less than last year's budget request, and \$251 million less than the final 1994 figure approved by Congress.

Proposed spending on the international space station remains at this year's level of \$2.1 billion, while space science rises very slightly (2.5 per cent) to \$1.77 billion. Funding for both aeronautics and human space flight would be reduced

from their 1994 levels.

Wesley Huntress, NASA's head of space science, says that this is "a good budget for us," with "no problems and no issues". Two large projects — the Cassini mission to Saturn and the Advanced X-Ray Astrophysics Facility (AXAF) — remain fully funded, as do two low-cost planetary projects, namely the MESUR Pathfinder Mars lander and the Near Earth Asteroid Rendezvous (NEAR) mission. In addition, NASA is asking for money for a continuing 'Mars Surveyor' programme, which would send spacecraft to Mars once every two years.

NASA's funding request for the Earth Observing System and its related data network jumps by 46 per cent to \$740 million as the project gets closer to its first launch in 1998, while funding for life sciences and microgravity research are reduced. The agency's Advanced Concepts and Technology office gets a healthy increase, from \$495 million to \$608 million. Of that amount, \$48 million is requested for advanced small satellite technology — four times this year's amount.

**T. R.**

# 'Error of judgement' over gene safety rules

**London.** Researchers from Birmingham University attracted widespread public criticism last week over their failure to meet the requirements of revised safety regulations on experiments with genetically engineered organisms. The criticism followed a decision by the Health and Safety Executive (HSE) to halt their research into the use of adenoviruses to insert oncogenes into human cells.

HSE officials acknowledge that the hazards of the work, being carried out in the university's department of cancer studies, are only "theoretical". But they acted swiftly to suspend the project after judging that the department had "failed to undertake a sufficient assessment of the risk to humans".

Their approach has been backed by John Beringer, professor of biology at the University of Bristol and head of the Advisory Committee on Releases to the Environment.

"Where there is any doubt, people have got to take extra care, and be seen to be taking extra care," says Beringer.

The HSE's action followed confusion over the level at which the research should have been classified. Under the Genetically Modified Organisms (Contained Use) Regulations that came into force in February last year, the HSE had to be notified of any level 2 work within a set period.

Members of the Birmingham research team, headed by Phillip Gallimore, thought they had fulfilled the necessary requirements by notifying the HSE of the precautions they had taken. But when an inspector visited the department last December, he decided that the research fell into a higher category of risk, and that the safety precautions being taken were therefore inadequate.

David Westbury, vice-principal of Birmingham University, says that scientists in

the department made "an error of judgement". But he adds that the affair forms part of a larger debate over the precautions needed to ensure the safety of experiments using genetically-engineered organisms.

The failure to meet the new standards had occurred partly because of the change in regulations, and partly because of an ongoing debate over the nature of the risks involved, he said. "It was a difference in opinion which is now being resolved."

But the HSE denies that the matter is merely a question of differing opinions. Working with the Advisory Committee on Genetic Modifications, the agency has produced detailed guidance on assessing the risks of particular types of experiment.

Two prohibition notices were served in December requiring the work to be temporarily halted, and an "improvement notice" was served last week specifying the safety changes needed if the work is to proceed.

The actions of the HSE have been welcomed by those concerned about the potential hazards of experiment using genetically engineered organisms as an indication that the agency's inspectors are prepared to use their powers to enforce the new regulations.

Professor Gordon McVie, scientific director of the Cancer Research Campaign, which is funding the work in question, says that the public "should be reassured that there is no danger that they could contract cancer as a result of the viruses used in these experiments".

But the publicity attached to the HSE's actions has not received such a warm reception. McVie acknowledges that the inspectorate has a useful role to play in "protecting scientists against false accusations"; but he adds that the publicity surrounding the affair has "not been the kind of publicity we want".

**Fiona Gammie**

## Research head backs students

**Munich.** Karl-Heinz Hoffmann, the new head of Germany's science council, the Wissenschaftsrat, has announced his support for students who are opposing a bill approved by the Bundestag last week that would freeze funding for student grants until 1996.

If passed into law, the new rules would require students to pass a progress test after two semesters in order to continue receiving a grant. The government's aim is both to save money and to encourage students to finish their courses more quickly.

Students throughout Germany held demonstrations and strikes last week against the rules. Despite the government's claims that many students spend too long at university, the students argue that they cannot be expected to finish their studies under pressure because they have to work to supplement their grants, which do not cover living costs.

Hoffmann, who took over as head of the Wissenschaftsrat from zoologist Gerhard Neuweiler two weeks ago, agrees with them. "Before we start to insist that study time is reduced, we must also make sure we can provide better conditions for students," he adds, referring to the overcrowded classrooms and poor-quality teaching found in most universities.

Hoffmann says he shares with his predecessor a deep concern for the next generation of German scientists. While agreeing that there are too many university students in general, he argues that there are "still too few scientifically qualified young people".

At least half of the science students at universities would be better placed in the more technically orientated higher education establishments, the *Fachhochschule*, he says. But the number of places must first

be increased; and more funding should be made available to young scientists once they are qualified, to help them to establish research careers.

Hoffmann, a 54-year-old professor from the Institute of Applied Mathematics in Munich, is keen to accelerate plans for improving the standards of teaching during his year as head of the science body.

Economic restrictions mean that the Wissenschaftsrat will be unable to achieve all its goals in the near future. But Hoffmann has certain priorities. First, he wants to address recent cuts to university library budgets which he considers unfair on students.

Second, Hoffman says he wants a rapid improvement in electronic communications networks between universities, research institutes and industry. The introduction of faster networks must be a priority for Germany, which lags behind other countries such as the US and Japan, he says.

Rainer Ortleb, the German minister of education, resigned last week after three years in office, shortly after the government approved his bill to save DM50 (US\$29.5 million) by reducing student grants.

The controversial bill has provoked widespread student action, and many of Germany's 16 state governments have already said they will vote against it in the upper house, the Bundesrat. Officials deny a connection between the bill and Ortleb's resignation, which they ascribe to health problems.

His place has been taken by Karl-Hans Laermann, a west German engineer and former university lecturer who was previously FDP spokesman on research policy and head of the party's committee on education, research and technology. **Alison Abbott**

## Rockefeller launches \$1 million challenge

**London.** In a bid to stimulate commercial interest in one of the major health problems of developing countries, the Rockefeller Foundation in New York has taken the unusual step of offering a prize of US\$1 million for the development of a simple kit to diagnose two sexually transmitted diseases (STD), chlamydia and gonorrhoea.

The scheme is open to anyone. But the foundation is hoping that the prize money will in particular attract biotechnology companies, which have experience of developing diagnostic kits. Such companies "do not really have an incentive for producing STD diagnostics for developing countries," says Seth Berkley, associate director of the foundation's health services division. □



# Nordic countries urged to boost internal cooperation

**Lund, Sweden.** Last month's entry into force of the European Economic Area (EEA) agreement between the 12 countries of the European Union (EU) and six countries of the European Free Trade Association (EFTA) means that all five Nordic countries (and Austria) are now able to participate fully in the EU's research programmes.

But a conference held here last week of the Nordic council of ministers (NCM) for education and research showed that their efforts to find a common political vision under these new circumstances are being hampered by economic difficulties, and by uncertainty over the adhesion of Sweden, Finland and Norway to the EU.

The Nordic countries — Denmark, Finland, Iceland, Norway and Sweden — are particularly concerned that the shift towards Europe may diminish inter-Nordic cooperation. "Are we going to continue [the NCM] or are we going to write it off?" asks Ole Vig Jensen, the Danish minister for education.

Nordic politicians have long sought closer regional cooperation, and indeed set an example by establishing freedom of movement between their countries in the 1950s. In the fields of science and education, the council has established a number of joint research institutes, postgraduate training schemes and scientific networks.

Furthermore, from the next academic year, students living in one Nordic country will be entitled to receive higher education in another country under conditions identical to those at home.

"The Nordic countries cannot assert themselves internationally if they do not work together," Per Unckel, the Swedish science



Northern enlightenment: Finland's Heureka science centre.

minister, told the meeting. He added that continued cooperation was particularly needed to boost the quality of Nordic research in order to improve the poor record of the region in attracting foreign researchers.

Another incentive to cooperation is that the Nordic countries are keen to preserve their egalitarian model of society, and indeed to export aspects of it to the EU. Antonio Ruberti, for example, the EU's research commissioner, told the conference that the EU could learn from their more advanced education and training systems.

But, as one observer at the meeting put it, "inter-Nordic cooperation has been based on negotiating different viewpoints, and we can't expect them to express a single viewpoint to Europe". National preoccupations may also increase because subscriptions to EU programmes will reduce the money available for both national and Nordic programmes.

Moreover, discussions about inter-Nordic cooperation in research have taken a back seat in several Nordic countries to the immediate problems caused by the end of decades of economic growth, and the switch from a welfare state to a free market philosophy (see *Nature* 360, 509; 1993).

For example, Riita Uosukainen, the Finnish minister for education, said that the economic difficulties faced by her country restricted her political vision to making cuts with the least damage. "There's not much else I can do."

Declan Butler

## Launch success marks Japan's space debut

**Tokyo.** Officials of Japan's National Space Development Agency gave a collective sigh of relief last week after the flawless launch of the H-II rocket, the first launch vehicle to be made entirely in Japan.

Development of the H-II's sophisticated first-stage engine has been plagued with problems and delays. But, apart from a few days' delay caused by bad weather and a minor technical problem, the launch went perfectly. The rocket placed two payloads in orbit, one being a re-entry vehicle that splashed down by parachute a few hours later in the Pacific Ocean, having orbited

the Earth before doing so.

The successful launch ends the agency's dependency on US rocket technology and opens the way for Japan to compete in the commercial launch market. After the launch, Satsuki Eda, the director-general of the Science and Technology Agency, denied that Japan's rocket development programme was intended to help turn the country into a major military power, and claimed that it would help turn space from a theatre for military competition into a sphere of activity "for people who want to give wings to peaceful dreams".

David Swinbanks

## UK newspaper goes quiet on challenge to HIV/AIDS link

**London.** Britain's *Sunday Times* appears to have backed off from its self-styled campaign to question the relationship between HIV and AIDS. So far this year, the newspaper has published only one feature article about AIDS (on 30 January), and this was a story that had already been reported extensively last year about a woman who is suing Wellcome over her husband's treatment with the antiretroviral drug AZT.

The recent lack of stories compares with the 25 — making an average of one story every two weeks — published by the newspaper during 1993. Most of these were intended to raise questions about the scientific validity of claims about the causal relationship between HIV and AIDS, and to cast doubt on the severity of AIDS epidemics in Africa.

The apparent cooling off in the *Sunday Times* campaigns follows a bold declaration of its "refusal to be silenced" on the issue in mid-December 1993. But by 6 January, readers wishing to join the debate were being sent letters saying that the correspondence was being closed.

Meanwhile, the position of the *Sunday Times* on AIDS in Africa continues to be challenged both by reports in the scientific literature documenting the continued spread of the disease, and by anecdotal reports from African countries (see *Nature* 367, 103; 1994).

For example, a large collaborative study from US and Ugandan researchers, published by the *British Medical Journal* in its issue of 15 January concludes that the incidence of HIV infection remains high in rural areas of Uganda, and argues that prevention strategies are urgently needed in the country.

The most recent journalistic investigation of claims by the *Sunday Times* newspaper that the spread of AIDS in Africa is a "myth" was a radio item broadcast by the BBC on 6 February as part of its midday current affairs programme *The World this Weekend*.

The programme featured an interview with Philippe Krynen, a French charity worker whose views form the central plank of the argument developed by the *Sunday Times* that there is no AIDS epidemic in Africa.

Krynen told the interviewer that he took pride in not being medically qualified, as this meant he was able to make up his own mind about the extent of AIDS in Africa. But his views contrasted sharply with those of a large number of doctors and other medical workers interviewed on the programme.

Maxine Clarke



# UK councils face £7.5 million 'efficiency' cuts

**London.** Britain's five research councils, shortly to be reorganized into six following the publication of last year's science white paper (policy document), have agreed to cut their administrative costs next year by £7.5 million (US\$10 million) to help ensure increased funding for industrially-related research projects.

William Waldegrave, the minister for science, said in a Parliamentary debate last week that the heads of the research councils had been persuaded to provide money from "efficiency savings" to a central fund.

The money being made available in this

way has been matched by a similar amount of "new money" provided by the Treasury. This will provide a total of £15.4 million, which will be redistributed to support research and postgraduate training that meet the goals of the white paper, namely focusing resources on fields expected to contribute to wealth creation.

According to Waldegrave, the reallocation of funding "augurs well for the increased efficiency that we are looking for." But representatives of the Institution of Professionals, Managers and Specialists, the union which represents many of those employed by the research councils, warns that job losses among administrative staff are "almost inevitable." Union officials point out that the planned "efficiency savings" coincide with an exercise launched by the government to discover which research council institutes are potential candidates for privatization, a move which they fear could also produce further staff reductions.

The biggest efficiency savings are being sought by the Science and Engineering Research Council, the bulk of whose activities will be divided from April 1 between two new councils, the Engineering and Physical Sciences Research Council (EPSRC), and the Particle Physics and Astronomy Research Council.

The EPSRC has agreed to find ways of

saving £3 million in administration costs, and PPARC £500,000. These figures compare with a total overhead expenditure by the SERC of about £25 million.

According to an SERC spokesman, some of the savings will result from the planned merger between the Rutherford Appleton Laboratory and the Daresbury Laboratories, and others from the introduction of a single management structure for the council's two UK observatories. But cuts are also likely to be made in support for the public understanding of science, and in staff training.

Nicholas Winterton, director of corporate affairs at the Medical Research Council, said that the MRC was planning to make further efficiency savings of £1.5 million on top of reductions already planned.

"We will be looking for further savings across the board," says Winterton. He adds that some of the MRC's expenditure plans for the next financial year may have to be postponed as a result, but that no reductions are expected to fall on research grants.

The least stoical of the research councils has been the Economic and Social Research Council, whose planned efficiency savings, although only totalling £200,000, still represents a significant proportion of overall administrative costs which have already been cut back significantly in recent years.

**David Dickson**

## Major: science part of 'back to basics'

**London.** Britain's Prime Minister, John Major, last week signalled a further step away from the science policies of his predecessor, Mrs Margaret (now Lady) Thatcher, by arguing that the government's support for science represents a key element of its "back to basics" campaign.

In the mid-1980s, Thatcher's officials justified cuts in the science budget on the grounds that science was a luxury that could be afforded only when a country became rich, and denied the need for a coherent policy for science and technology.

But Major told the annual luncheon of the Parliamentary and Scientific Committee — a body made up of both parliamentarians with an interest in science and research leaders from the academic and industrial community — that making the best use of the nation's scientific abilities was a "common-sense, basic building-block for civilized living".

The "back to basics" campaign is designed to focus political support on social priorities. Major said that, in preparing for last year's white paper (policy document) on the organization of science, the government had been "clearly right" to question its previous role.

He pointed out that, despite the need to restrain public expenditure, spending on the science base was expected to rise in real terms next year. Major also promised that "science will remain a high priority in future".

Major's remarks were welcomed by the group Save British Science (SBS), which said that it had "come a long way" since pleas for greater political support for science fell on deaf ears in the mid-1980s. But John Mulvey, the secretary of SBS, says that funding prospects remain fragile. "If the positive commitments made are, as John Major says, aspects of his 'back to basics' theme, then we need more 'basics' — and fast," says Mulvey. □

## Waldegrave announces new awards

**London.** A new award scheme with the somewhat unattractive name of ROPA — named after last year's science white paper (policy document) *Realizing our Potential* — was launched last week by Britain's minister for science, William Waldegrave, in a move designed to stimulate research in areas promising to contribute to wealth creation.

Awards under the new scheme, which will be run jointly by three research councils, will be made to scientists who wish to work on long-term projects with potential industrial applications, and are already receiving support from industry.

The government has promised to provide £3.5 million (US\$5.2 million) in the financial year beginning on 1 April for a six-month pilot scheme for the new awards. It is one of a package of initiatives totalling £15.4 million which the government has decided to fund in addition to previous plans for the research councils.

In line with the priorities outlined in the white paper, the government is to provide an extra £2 million to an 'innovative manufacturing' programme to be run by the new Engineering and Physical Sciences Research Council, £3.4 million to the human and animal genome and immunology pro-

grammes of the Medical Research Council and £4 million for chemistry research.

It has agreed to allocate an additional £600,000 to the CASE award scheme, designed to back postgraduate students carrying out research projects partly in an industrial environment, and also to increase the number of fellowships offered to young scientists through the Royal Society's University Research Fellowships Scheme.

But the government has turned down a request from the Economic and Social Research Council (ESRC) to provide continued funding for a programme of research into risk and behaviour. This move prompted the chairman of the ESRC, Howard Newby, to break ranks with his fellow chairmen by announcing his "disappointment" with the government's deliberations on how the new money should be distributed.

Announcing the details of the new funding in a parliamentary debate, Waldegrave said that a unique feature of the ROPAs awards is that industry's recognition of researchers — rather than traditional peer review — would be used as an indicator of quality and relevance, giving industry a "major role" in identifying which researchers should be supported. □

## NIH panel opens debate on ethics of embryo research

**Washington.** An advisory panel convened by the US National Institutes of Health (NIH) to look at the moral and ethical issues raised by the use of human embryos in federally funded research got off to a slow but steady start last week.

The panel has been set up to help Harold Varmus, the director of NIH, to develop guidelines governing the review and conduct of such research. It has less than six months to prepare a report on its recommendations, which will be presented at the July meeting of the advisory committee to the director of NIH.

Last June, Congress lifted the *de facto* ban on the use of federal funds for human embryo research. Since then, NIH has received almost 40 grant proposals for support in this and the related area of parthenogenesis. Although the proposals fall outside the jurisdiction of current guidelines, all have received approval by local institutional review boards.

The 19-member panel consists of experts in basic and clinical research, ethics, law, social science, public health and public policy and is chaired by Steven Muller, president emeritus of Johns Hopkins University, in Baltimore, Maryland.

Varmus has asked the panel members to advise him on which areas of human embryo research are deemed acceptable, which warrant further review, and which are unacceptable. The highly controversial area of germline gene modification is not within the purview of this panel.

Before last June, research involving *in*

*vitro* fertilization (IVF) had to be reviewed by an ethics advisory board (EAB). In a report issued in 1979, the board found it ethically acceptable for the federal government to support or conduct research involving human IVF and embryo transfer. But the dissolution of the board in 1980 has effectively blocked federal funding of such research for the past 13 years or so.

Despite the regulatory blockade, research — most of which has been directed towards improving the outcome of IVF — has been undertaken in the private sector. But Jonathan Van Blerkom, of the department of molecular, cellular and developmental biology at the University of Colorado, said that much of it has not been “meaningful”, claiming that many studies were poorly designed.

Varmus told the panel that although the EAB report was timely in 1979, “the passage of 15 years, and a number of scientific and social changes, limit its applicability”.

Charles McCarthy, former staff director of the EAB and now a member of the board of Public Responsibility in Medicine and Research, said that seeing the issues finally under discussion was “a dream come true”.

Over the next six months, the panel will address a number of controversial issues surrounding human embryo research, such as whether it is permissible to create embryos solely for research purposes rather than use ‘spare’ embryos created during IVF, the use of fetal ovaries as a potential source of ova with which to create embryos, and embryo ‘cloning’ (twinning).

**Diane Gershon**

## France seeks views on improvements in research system

**Paris.** Sixty thousand copies of a report on the issues facing French science were distributed to laboratories last week as the government launched its national consultation on research. The exercise has been planned as a run-up to a debate in the National Assembly in June which the government plans to use to launch a new national strategy for research.

The report sets the agenda for six regional meetings being held over the coming weeks in Marseilles, Grenoble, Bordeaux, Strasbourg, Le Mans and Lille. But it will disappoint those impatient for new ideas. Its main content is a long list of deficiencies of the research system, many already well-known. Members of the committee that drafted the report in consultation with scientific and industrial organizations and trade unions say they avoided specific recommendations in order to leave the debate open.

“Lack of time” is blamed for the report’s admission that it is not exhaustive. But it omits nuclear research, despite its large share of the civil budget — ostensibly because the National Assembly will discuss this as part of a debate on energy.

Similarly, the report omits any consideration of defence research, even though the government seems increasingly committed to pursuing an integrated approach to civilian and military research (see *Nature* 367, 102; 1994).

Many of the deficiencies described in the report result from what François Fillon, the minister for higher education and research, has called the “worrying rigidity” of the French system. In particular, it criticizes the fact that salaries and overheads lock up most of the budgets of the research organizations, leaving them little money for research.

The report also suggests that France should revise the goals of its costly *grands programmes technologiques* in aeronautics, space, nuclear science and computing. In particular, says the report, France should “state clearly” whether it needs a manned space programme or a new generation of nuclear reactors.

More than 1,800 scientists and industrialists are expected to attend the planned regional meetings, and a further 1,500 a national summing-up in Paris on 9 April. Edouard Balladur, the prime minister, will take part in some of the meetings.

The government has also established a special electronic bulletin board, postbox and fax line to let other researchers make their views known. But whether it intends to heed what is said — or simply, as some fear, to use the consultation to claim a mandate for a *fait accompli* — remains to be seen.

**Declan Butler**

## Japan sets up gene therapy panel

**Tokyo.** Japan’s Ministry of Health and Welfare announced on Monday (7 February) the members of a committee that will screen all proposals for investigating the clinical applications of gene therapy, opening the way for research in this field in Japan.

But the Ministry of Education, Science and Culture has yet to make clear what regulatory system it will establish for screening such research in universities.

The 17-member committee is made up of specialists from a broad range of fields. They include: Kenichi Matsubara, head of Japan’s human genome project; Masaki Terada, director of the National Cancer Centre Research Institute; Ayako Sono, a well-known novelist and Suzuo Matsuda, a former journalist with Jiji wire service.

All applications for experiments involving gene transfer in humans will be examined by the committee, including those to be

carried out in universities, after initial screening by an internal committee within the relevant institution.

The new committee will therefore function in a similar fashion to the Recombinant DNA Advisory Committee of the US National Institutes of Health. But, unlike the US committee, it is not expected to open its deliberations to the press and public; government committees in Japan usually carry out their deliberations behind closed doors.

The education ministry has yet to clarify its own position, and has in fact been approaching members of the ministry of health committee, asking for their help in establishing a separate screening system for the universities they oversee. But this is causing confusion among university researchers who are uncertain about where they must go to obtain approval for proposed experiments.

**David Swinbanks**



# Hungary's academy fights to build a new role

**Budapest.** A new law redefining the responsibilities of the Hungarian Academy of Sciences as it steers basic research into the post-communist era is still awaiting approval by the parliament, nearly three years after it was introduced.

The bill would give legal backing to the academy's status as a politically independent learned society with a more democratic structure than in the past, and with control over its 39 research institutes. The academy wants this status confirmed as soon as possible to help it negotiate for more funds; its budget has been cut so drastically in recent years that research is almost at a standstill.

But politicians remain divided on the academy's future, partly because many want its power to be reduced, and partly because basic science is not a priority in Hungary's troubled economy. As a result, it now seems unlikely that the bill will be accepted before elections due in May.

The academy was founded as a learned society in 1825. In the 1920s and 1930s, it set up the country's first independent research institutes. But extensive reorganization took place along the lines of the Soviet Academy of Sciences after the communists came to power in 1949.

At the time, 122 members of the academy were expelled, its political autonomy was abolished — the academy acted as a ministry of science, with control over the allocation of public funds for basic research — and an extensive network of research institutes was established.

Since the election of a post-communist government in 1990, the academy has not been represented in the government, and its grant-giving section, the Hungarian Science Research Fund (OTKA) has been made independent.

The academy has been quietly introducing its own democratic reforms. For example, it is creating two different levels of scientific board, which include scientists from universities and industry, but exclude institute directors, to decide how money should be divided between its institutes. But political problems remain.

During the four decades of communist rule, for example, the academy provided a research environment that was much richer and freer than that of the universities. It also offered shelter for scientists considered 'politically unfit' for university teaching posts.

Despite this, it now finds that it is having to defend itself against charges that it is perpetuating the influence of communism,

despite the election of the anti-communist Domokos Kosáry as president.

Kosáry, a well-known historian who spent four years in jail as a dissident, immediately proclaimed that there would be no 'witch-hunt' against former communists in the academy. But this angered the more zealous reformers in parliament.

He also immediately rehabilitated all of the academy members expelled by the communists in the early 1950s apart from one who was judged to have actively helped the fascists during the Second World War.

But this move has also backfired. Two weeks ago a small far-right party, the Foundation of Victims of Communism, rekindled arguments about whether communists still hold power in the academy by claiming that this exception was made because of the scientist's anti-Stalinist politics. The academy vigorously denied the allegation, but in doing so displayed its hypersensitivity to charges of communist sympathies.

Controversy over the academy bill is also being fuelled by Hungary's economic difficulties. Industrial output has been falling steadily since 1989, and many feel that the country cannot afford the academy's demand to run its research institutes along the lines of Germany's Max Planck Society, where the state foots the bill but has no influence over the direction of research.

Many politicians favour spending any money available on short-term applied research to help Hungary out of its current crisis. Others believe that the research institutes should be handed over to universities.

But scientists in the research institutes wish to remain under the academy's umbrella, says Pál Venetianer, director of the academy's Biological Research Centre in Szeged, south Hungary. They have welcomed the academy's internal reforms, and believe it is now well placed to protect their interests.

The academy is working hard to protect the future of its institutes. But this means negotiating realistic funding, which in turn requires an accepted legal status. "The finance minister does not take us seriously at the moment," says László Keviczky, the academy's secretary general.

The academy's annual budget has fallen sharply, from 5.15 billion Hungarian forint in 1992 to Ft4.86 million (US\$49 million) in 1994, at a time when inflation has been running at between 20 and 30 per cent. "We barely have enough money to cover running

costs of our institutes," says Keviczky.

Project money distributed by OTKA has also fallen dramatically; this year for example saw a cut of more than 20 per cent, from Ft2.44 billion to Ft1.92 billion. The academy would like to be able once again to provide project money of its own, says Kosáry, something he will fight for once the academy law is passed.

Between 1990 and 1992, the academy organized international evaluations of its institutes. A common criticism was that institutes were overstaffed yet individual research groups were too small. Staff numbers have fallen by nearly a third, from 7,580 in 1990 to 5,440 now. But this has mostly been the result of scientists leaving the country or going into industry, and by the retirement of many ageing staff.

The main problem now facing the academy is how to keep the remaining staff during the current period of uncertainty and economic difficulties. The average wage of a scientist at group leader level is only US\$300 a month, a quarter of the equivalent salary in industry. A postdoc earns between US\$180 and \$200, a research director, US\$400. Even the 250 academy members have seen their previous privileged salaries reduced to US\$850 a month.

Low wages and scarce research funds are taking their toll of scientists, many of whom have spent long periods working in the West and are frustrated by the near-impossibility of their situation at home. "There has never been more pressure on us," says Pál Ormos, a 40-year-old director of research at the Biological Research Centre in Szeged. He is, however, determined to ride out the storm.

**Alison Abbott**



**Kosáry: healing old wounds.**



**Venetianer: pleased with reforms.**

## Howard Hughes opens up awards selection

**Washington.** Forty-four researchers — 17 at the assistant professor level — are the winners of a nationwide competition run by the Howard Hughes Medical Institute (HHMI) aimed at searching out the best and the brightest that biomedical research in the United States has to offer.

The "competition", made possible by an HHMI endowment that increased by US\$780 million during 1993, provided the institute with an opportunity to open up the selection and appointment process beyond the 53 universities and medical centres that have HHMI investigators. Indeed, ten of the new appointments fall into this category.

"The best way to make a large number of appointments is to solicit nominations broadly and have them all reviewed together," says Purnell W. Choppin, president of HHMI. □



## AIDS in Africa: is it a myth?

**SIR** — In *The Sunday Times* of 21 March 1993 in an article entitled "Epidemic of AIDS in Africa 'a tragic myth'" the following appears:

A British expert puzzled by the disparity between apocalyptic warnings about the spread of AIDS in Africa and her own observations is Professor Beverley Griffin . . .

For the past seven years she has received blood samples of hundreds of children in Malawi. The proportion that are HIV-positive has remained unchanged, at 1–2%. This casts doubt, she feels, on claims made elsewhere that Malawi is in the grip of an HIV epidemic, with about a fifth of its population infected.

Our letter in response to this article (*Nature* 366: 716; 1993) argued that such prevalences amongst children reflect a high prevalence among adults. We fail to see how commenting on data published in a national newspaper is a breach of ethics and etiquette as Griffin claims (*Nature* 367, 212; 1994). We suggest that Griffin should take up any grievances with the author of the article in *The Sunday Times*.

We also fail to see how the prevalences quoted by Griffin from our paper (R. M. Anderson *et al.* *Nature* 352, 582–589; 1991) illuminate the debate. We clearly point out that the populations in question were urban, and repeat this in our letter responding to *The Sunday Times* article.

**Geoff Garnett**

**Roy Anderson**

*University of Oxford,*

*Department of Zoology,*

*Infectious Disease Epidemiology Unit,*

*Oxford OX1 3PS, UK*

## When to speak

**SIR** — In your leading article (*Nature* 367, 301; 1993) about the prison sentences served by Dr Michel Garretta and Professor Jean-Pierre Allain, you ask what a public official is to do if he knows official policy to be dangerous. "To broadcast his disquiet to the public, or to attempt to work within the bureaucracy that employs him?" In fact, Allain did give an interview on his fears to the French newspaper *Le Matin*, but it was not published. His wife, the immunologist Dr Helen Lee, tried to warn the public at a meeting of 200 leading haematologists and other experts and was officially reprimanded for doing so. As I understand it, a recent ruling by our own government expressly forbid civil servants to take such a step on pain of disciplinary action.

**M. F. Perutz**

*MRC Laboratory of Molecular Biology,  
Hills Road, Cambridge CB2 2QH, UK*

## Biased view

**SIR** — Our companies publish three competing bibliographic software products: Reference Manager, EndNote and ProCite. On 11 November, *Nature* published a review of another bibliographic software package (*Nature* 366, 183–84; 1993). The article was written by the producers of the product being reviewed and they compared their product to ours. We feel that it is unethical and misleading to allow authors to review their own software. As the article appeared in your Product Review section, readers may perceive it as an unbiased review, which it is not. We appreciate accurate critical reviews that give the reader a basis on which to choose a product, but a 'review' by the author of a work can only be biased.

**Victor Rosenberg**

*Personal Bibliographic Software, Inc.,*

*PO Box 4250, Ann Arbor,*

*Michigan 48106-4250, USA*

**Richard Niles**

*Niles & Associates, Inc.,*

*2000 Hearst Avenue, Suite 200,*

*Berkeley, California 94709, USA*

**Earl B. Beutler**

*Research Information Systems,*

*2355 Camino Vida Roble,*

*Carlsbad, California 92009, USA*

## Nuclear Sweden

**SIR** — In 1980, the Swedish population decided in a referendum to phase out nuclear power by 2010. This choice hinged on being given the time to develop suitable alternatives. Today, the Swedish government seems to be hesitating about fulfilling its obligation, and nuclear plant operators and organizations such as the Swedish Royal Academy of Engineering Sciences (IVA) are openly advocating that the referendum decision be reversed.

The argument is that time has shown that, despite huge efforts, it has not been possible to provide a viable alternative to nuclear power<sup>1</sup>. Although the Swedish government has spent quite large sums on exploring alternatives, much of the money was spent on paper studies, including a stream of alternative energy plans. Industrial efforts to develop alternatives have not been successful in Sweden. Yet in neighbouring Denmark, a wind turbine industry has grown into a billion-kroner export business. And solar thermal collectors seem to be more abundant in most other European countries than in Sweden.

Swedish government policy has ensured that no alternatives were given a chance. No environmental tax has ever been levied on cheap Swedish electricity, and no market for wind power, for example, has been created. In Denmark, a 30 per cent subsidy helped the wind industry to

get started. This subsidy was given to customers, and a requirement that the full investment should be recovered in less than 8 years prevented manufacturers from including the subsidy in their pricing. Subsequently, as wind energy became increasingly viable, the subsidy was reduced to 20 per cent then to 10 per cent, and then abandoned.

Sweden uses all its hydro and nuclear power generating capacity, and has to import electricity every time the nuclear plants are closed with some problem (which recently happened to all of them at the same time). Sweden has no alternative to nuclear power because it deliberately chose a policy barring the development of such alternatives.

**Bent Sørensen**

*Roskilde University, Institute 2,*

*PO Box 260, DK-4000 Roskilde, Denmark*

1. Engström, G. *Energy for coming Generations*. In *Proc. 5th Int. Energy Conference*, 1993, Seoul, South Korea (in the press).

## Let's have a word

**SIR** — Work has recently begun at Oxford University Press on a comprehensive revision of the *Oxford English Dictionary* (*OED*), to be published in 2005.

The published literature of science is so vast that it is impossible to scan all of it for relevant information. As part of the process of revising and updating the dictionary, we should like scientists who have relevant information to contact us.

As an example, Professor Joshua Lederberg e-mailed us about the omission of C. S. Peirce's philosophical sense of 'abduction', not recorded in the *OED*, and his own coinage of the words 'eugram' and 'eugraphy'.

Information of particular interest would include: the coinages of particular scientific words, where this information is not already given; factual errors in definitions; scientific words and meanings not in the *OED* (second edition, 1989); and referenced examples of words and meanings that predate the earliest quotation given in the *OED*.

The last category is especially useful: the *OED* is a historical dictionary that aims to trace every word and meaning back to its earliest known use in the English language, as well as giving references to the coinages of foreign precursors of English words. Such first uses can be difficult to track down; for example, we have been unable to find any use of the word 'multiplicatively' earlier than 1914, even though it is in Funk and Wagnalls' *Standard Dictionary* of 1895.

**A. M. Hughes**

*(Chief Science Editor)*

*Oxford English Dictionary,*

*Oxford University Press,*

*Walton Street, Oxford OX2 6DP, UK*

e-mail: ahughes@oup.co.uk.

# Better protection of the ozone layer

Malcolm K. W. Ko, Nien-Dak Sze & Michael J. Prather

**How can we extend the Montreal Protocol to other ozone-depleting chemicals, such as fuel from the Space Shuttle and pharmaceuticals, when the life cycles of these compounds and the scales of the industries are different?**

INTERNATIONAL agreements have been enacted to protect the ozone layer by regulating the release of chlorine- and bromine-bearing chemicals such as the chlorofluorocarbons and the halons. One of the criteria by which such chemicals are assessed and regulated is the ozone depletion potential (ODP)<sup>1-5</sup>. But because more and more chemicals are turning out to be ozone-depleting, we believe that a more refined approach is needed for effective and equitable control. Here, we address ways to extend international agreements to include other contemporary ozone-depleting chemicals whose applications and life cycles are very different from synthetic halocarbons used in the industries already singled out for concern, such as refrigeration, insulation and fire-fighting applications.

During the past decade, an international regime has emerged for limiting the consumption of chemicals believed to carry chlorine (Cl) and bromine (Br) into the upper atmosphere (stratosphere) (see box). The Convention on the Protection of the Ozone Layer (Vienna, 1985) was followed in 1987 by the Montreal Protocol, subsequently amended in 1990 (London amendment) and in 1992 (Copenhagen amendment). The governments that have signed these treaties are committed to phase out the production and consumption of CFCs (chlorofluorocarbons) and halons (brominated hydrocarbons), and to limit the production of their substitutes (for example hydrochlorofluorocarbons, HCFCs) in the expectation that the stratospheric concentration of chlorine and bromine can thereby be reduced. Ultimately, the goal is to reverse the observed downward trend of global ozone, and to limit the possible consequential damage to the biosphere from increased ultraviolet radiation. The adherence of governments to these agreements, which have the status in international law of formal treaties, is mirrored in national legislation such as the US Clean Air Act (1992).

The evolving protocols have given precedence to controlling synthetic CFCs and halons used in applications such as electronic and metal cleaning, foam blowing and refrigeration (CFCs); and fire extinguishers (halons). These chemicals are released in the lower atmosphere (the troposphere), but are sufficiently long-lived that they can be transported to the stratosphere where most of them are

broken down by ultraviolet radiation, producing highly reactive Cl and Br radicals that are chiefly responsible for the catalytic destruction of stratospheric ozone. Because the timescale of mixing in the troposphere is less than the residence time of these halocarbons, the effect on ozone (as measured by the ODP) does not depend exactly on where, when and how they are released.

But there are other more direct and effective means by which chlorine can enter the stratosphere. These include solid-fuel rocket motors in the Space Shuttle launches, which deposit chlorine directly in the stratosphere. Prudence, as well as consistency, requires that these sources should also be evaluated under the same criteria (for example, ODPs) to determine their contributions, if any, to ozone depletion. Before the Copenhagen meeting, the scientific community was asked to quantify the impact of other atmospheric emissions such as solid rocket

motors<sup>6</sup>, stratospheric aircraft<sup>7</sup> and use of methyl bromide (CH<sub>3</sub>Br)<sup>8</sup> in agricultural activities. But no ODPs were calculated for solid rocket motors and stratospheric aircraft. Based on the findings on CH<sub>3</sub>Br, a freeze on its production beginning in 1995 was adopted in the Copenhagen amendment.

What follows is a demonstration that some potential ozone-depleting substances have life cycles that differ qualitatively as well as quantitatively from the chemicals now controlled by the agreements. We discuss what factors should be considered when developing a strategy for control of compounds whose applications and life cycles are very different from the CFCs. Specifically, we look for ways to ensure that the resulting strategy is practical (can it be applied easily?), effective (does it omit chemicals whose net impact may be comparable to individual HCFCs or CFCs?), and equitable (does it ban or impose excessive penalty on uses that

## Chronology of ozone-protection agreements

THE Vienna Convention for the Protection of the Ozone Layer was agreed in March 1985 and entered into force in September 1988. It was set up with the intention of preventing any further damage to the ozone layer by "recognising the possibility that world-wide emissions and use of fully halogenated chlorofluorocarbons (CFCs) and other chlorine-containing substances can significantly deplete and otherwise modify the ozone layer, leading to potentially adverse effects on human health, crops, marine life, materials and climate, and recognising at the same time the need to further assess possible modifications and their potentially adverse effects."

It was also agreed that negotiations would continue on the development of a protocol to control equitably global production, emissions and use of CFCs. From this, the Montreal Protocol, an international agreement to phase out ozone-depleting substances, was agreed in 1987.

■ 1987. Montreal Protocol on Substances that Deplete the Ozone Layer. It was agreed that global action was needed to phase out CFCs and halons.

■ 1989. First meeting of the parties to the Protocol in Helsinki. Declaration

adopted calling for CFCs and halons to be phased out by 2000.

■ 1990. Second meeting of the parties to the Protocol in London. Agreement to phase out CFCs, halons and carbon tetrachloride by 2000 and methyl chloroform by 2005. Financial mechanism set up to assist developing countries.

■ 1991. New European regulation 594/91 came into force. CFCs to be phased out within the European Communities by 1997.

Second meeting of the parties to the Vienna Convention and third meeting of the parties to the Montreal Protocol.

■ 1992. Fourth meeting of the parties to the Montreal Protocol in Copenhagen. Parties agreed to bring forward phase-out dates for CFCs, carbon tetrachloride and methyl chloroform to 1996; halons to be phased out by 1994. Controls agreed for methyl bromide and HCFCs.

■ 1993. Fifth meeting of the parties to the Montreal Protocol. It was agreed that there would be no essential uses for halons in 1994. Multilateral fund replenished.

Third meeting of the parties to the Vienna Convention. □

Source: UK Department of the Environment.



have relatively small effects on the ozone layer?). We will use three examples to illustrate these points: chlorine deposited by the solid-fuel rockets from the Space Shuttle launches;  $\text{CH}_3\text{Br}$  used in soil fumigation; and brominated compounds in pharmaceutical use.

### Ozone depletion potentials

First defined for CFCs a decade ago, the ODP is an index measuring the time-integrated ozone depletion caused by specific quantity of a chemical relative to that caused by the same quantity of the chlorofluorocarbon CFC-11 (the fully substituted methane  $\text{CFCl}_3$ ). The definition presumes the chemical is ultimately released into the atmosphere. Policy makers and industry representatives have since asked scientists to extend and calculate ODPs for HCFCs<sup>2</sup> and the halons<sup>3,5</sup>. Total chlorine loading of the atmosphere<sup>9</sup> has also been used to assess the global ozone loss caused by these chemicals, either separately or in combination for specific emission predictions. The amount of chlorine in the stratosphere not still tied up in the parent halocarbon is defined as the stratospheric chlorine loading<sup>10</sup>. To understand the relation between ODP and chlorine loading, it is necessary to examine the life cycles of the halocarbons.

The life cycles of the halocarbons are illustrated in the figure. These halocarbons are, in general, relatively long-lived and thus well mixed in the troposphere. They cycle between the troposphere and stratosphere where a portion of the compound is dissociated during repeated excursions into the stratosphere and the troposphere. The Cl and Br atoms formed in the stratosphere participate in catalytic cycles removing ozone until, on average after 3 years, they are transported back to the troposphere. The Cl and Br atoms in the troposphere, either produced *in situ* by degradation of the halocarbons or transported from the stratosphere, are assumed to be removed quickly by rain out or surface deposition. For CFCs, almost all the chlorine atoms are produced in the stratosphere. The HCFCs react with hydroxyl radicals (OH) and thus are destroyed readily (on a timescale of a few years to decades) in the troposphere. As a result, they contribute less to the stratosphere chlorine loading on a per-molecule-emitted basis.

Numerical simulations using photochemical models indicate that the time-integrated stratospheric chlorine loading is a good proxy for cumulative ozone loss<sup>9,10</sup>. We have used a simple two-box model (with tropospheric and stratospheric compartments) to calculate, as a function of time after release, the amounts of chlorine still tied up in the undissociated halocarbons in the troposphere ( $B_T$ ) and stratosphere ( $B_S$ ), as well as the quantity of free chlorine in the stratosphere ( $C$ ,

stratospheric chlorine loading). The differential equations describing the evolution over time of the three reservoirs are

$$\begin{aligned}\frac{dB_T}{dt} &= -\frac{B_T}{L_T} - \frac{B_T f - B_S}{\tau_t} \\ \frac{dB_S}{dt} &= -\frac{B_S}{L_S} - \frac{B_S - B_T f}{\tau_t} \\ \frac{dC}{dt} &= -\frac{C}{\tau_t} + \frac{B_S}{L_S}\end{aligned}\quad (1)$$

where  $L_T$  and  $L_S$  are the tropospheric and stratospheric lifetimes for chemical loss of the halocarbons and  $\tau_t$  is the turnover time for replacing stratospheric air by tropospheric air. In calculating the exchange flux, the tropospheric burden is scaled by a factor  $f = 0.15/0.85$  (assuming here that 15% of the atmospheric mass is in the stratosphere) so that the flux is proportional to difference in the mixing ratios. We select  $\tau_t = 3$  yr, consistent with Holton's<sup>11</sup> derivation of 2.5 yr for the turnover of air above 100 mbar. Based on two-dimensional model results<sup>12</sup>, we select  $L_T = 1,000$  yr and  $L_S = 5$  yr for CFC-11.

Assuming  $B_T(0) = 1$  kg chlorine in the form of CFC-11,  $B_S(0) = C_{F-11}(0) = 0$ , the burdens (in kg chlorine) for year  $t$  are

$$B_T(t) = 0.930 \exp[-t/46.5] + 0.070 \exp[-t/1.75] \quad (2a)$$

$$B_S(t) = 0.107 (\exp[-t/46.5] - \exp[-t/1.75]) \quad (2b)$$

$$C_{F-11}(t) = 0.068 \exp[-t/46.5] + 0.090 \exp[-t/1.75] - 0.158 \exp[-t/3] \quad (2c)$$

The integrated chlorine loading (p.p.t.v.-year) for emission of 1 kton ( $1 \times 10^6$  kg) of CFC-11 is obtained by integrating equation (2c) and multiplying by 0.81.

$$\begin{aligned}IC_{F-11}(t) &= 2.33 (1 + 0.16 \exp[-t/3] - \\ &\quad 1.11 \exp[-t/46.5] - \\ &\quad 0.05 \exp[-t/1.75]) \text{ p.p.t.v.-year}\end{aligned}\quad (3)$$

The ODP of a compound  $X$  can be approximated as the ratio  $IC_X(t)/IC_{F-11}(t)$ , where  $IC_X(t)$  is the integrated chlorine loading due to 1 kiloton release of  $X$  calculated using appropriate  $L_T$  and  $L_S$ . The steady state ODP corresponds to the ratio in the limit as  $t \rightarrow \infty$ .

### Regulating halocarbons

The Protocol to date has focused on limiting the production of halocarbons used in traditional applications. The list of products containing controlled species (annex D of the Protocol) specifies automobile and truck air-conditioning units; domestic and commercial refrigeration

equipment; aerosol products (except medical aerosols); portable fire extinguishers; insulation boards, panels and pipe covers; and pre-polymers. The list of controlled species (annexes A-C) is mainly limited to synthetic fluorochlorocarbons or bromocarbons with no known natural sources.

Although the Protocol recognizes the ODP as a measure of the environmental danger, the lists of controlled species in the annex are not grouped by ODP values. Rather, the species were classified according to chemical type, by use and by when they were included in the Protocol. However, the ODP values are used explicitly in the following way. Instead of imposing production limits on an individual chemical, the control is applied to groups of chemicals. Within each group, the production of each chemical can be traded after appropriate weighting by its ODP value. The Protocol states that "each Party shall, for each group of substances ... determine its calculated levels of Production by (i) multiplying its annual production of each controlled substance by the ozone depleting potential specified ...; (ii) adding together, for each such Group, the resulting figures." Under the Copenhagen amendment, the annual production limit on the transition substances (HCFCs) are specified as the 1989 production plus 3.1% of the ODP weighted annual production of the CFCs in the same year.

Although it is not explicitly stated, a reading of the protocols and the US Clean Air Act suggests that their formulations are based on considerations of both ODP and chlorine loading from the current emission rates. There is agreement to phase out production of the CFCs (annex A group I and annex B group I) and  $\text{CCl}_4$  (annex B group II) by 1996; and halons (annex A group II) by this year. These compounds all have ODPs of about 0.5 or larger. The atmospheric ( $1/e$  folding) lifetimes of these compounds are 30 years or longer (with the exception of halon-1211). Once introduced into the atmosphere, these chemicals will continue to release Cl/Br atoms into the stratosphere over their respective lifetimes. Thus, even if their production is stopped immediately, it will take many decades (until the chemicals are purged from the atmosphere) before the chlorine concentration can return to the level of about 2,000 p.p.t.v. which existed before the ozone hole<sup>10</sup>.

In the Copenhagen (1992) amendment, there is a proposal to phase out the substitute HCFCs (annex C) by the year 2030, and to impose stringent limitations on production rates during the interim period. The HCFCs, with atmospheric lifetimes of about 15 years or less, are more rapidly purged from the atmosphere and thus allow for more rapid decay of stratospheric chlorine once use of the

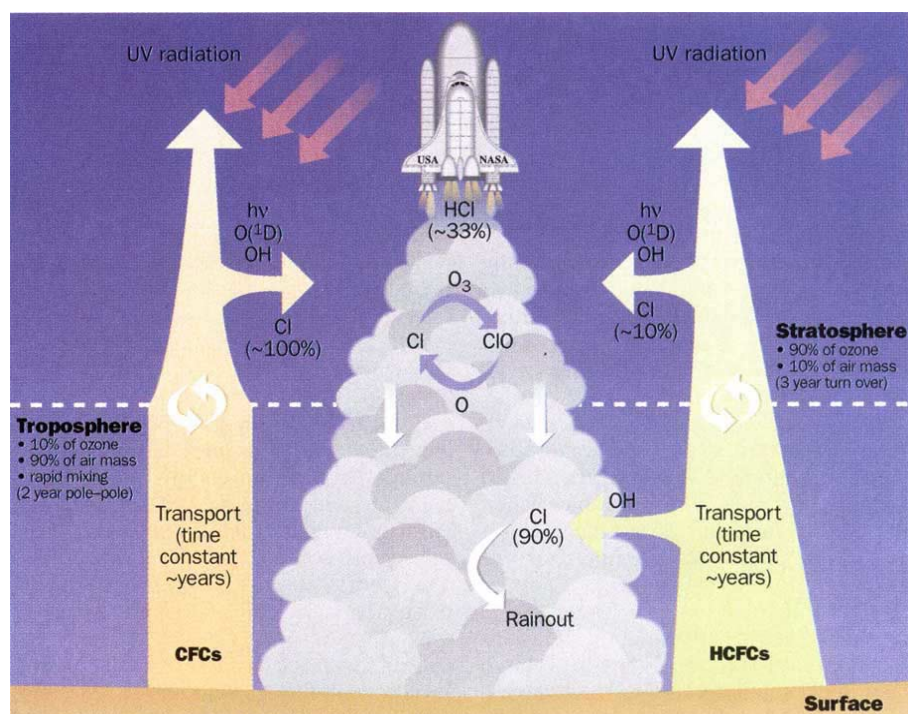
chemicals is phased out. One rationale for controlling HCFCs with ODPs as small as 0.02 is that the short-term relative impact of such species is much larger than suggested by the steady-state ODP<sup>4</sup>.

Methyl chloroform ( $\text{CH}_3\text{CCl}_3$ , annex B group III), a compound that fits loosely with the definition of HCFC (but without fluorine), has an atmospheric lifetime of about 6 years and an ODP of 0.12. It is due to be phased out in 1996. This decision was presumably made because the very large production rate contributes significantly to the chlorine loading of the current atmosphere. Chemicals in another class — methylene chloride ( $\text{CH}_2\text{Cl}_2$ ), perchloroethylene ( $\text{C}_2\text{Cl}_4$ ) and trichloroethylene ( $\text{C}_2\text{HCl}_3$ ) — have annual productions comparable to that of  $\text{CH}_3\text{CCl}_3$ , but have not yet been regulated, presumably because they have much shorter lifetimes, smaller ODPs and give rise to smaller chlorine loading.

Some countries have formulated unilateral regulations based mainly on the calculated values of ODPs. Under the US Clean Air Act, any chemical with an ODP larger than 0.2 can be classified as a class I substance. Unless they are in current production, such chemicals may not be manufactured. Heavy fines (up to \$25,000 per violation per day) can be imposed for any intentional venting of these materials; and indeed the Environmental Protection Agency is currently seeking large fines from 28 individuals and businesses for such violations<sup>13</sup>. This ODP cut-off is applied independently of the amount released, the argument being that the combined effects from a large variety of ozone-depleting chemicals, even though produced in small quantities, will add up and must be regulated as a collective industry. The ODP cutoff is also applied regardless of how the chemicals will be used. As illustrated by the examples that follow, this approach may be problematic when extended to control chemicals outside the traditional CFC industries because the ODP may depend sensitively on end-use, which in turn determines the life cycle or fate of the chemicals.

### ODP of solid rocket fuel

Most scientific studies and early regulations have focused on chemicals whose atmospheric life cycles and participation in ozone depletion are similar to that of CFC-11. The chlorine loading from Space Shuttle launches requires a somewhat different approach. The solid rocket fuel is composed of 16% by weight of aluminum, 70% of ammonium perchlorate ( $\text{NH}_4\text{ClO}_4$ ), and 14% of a polymer matrix. Combustion gives rise to the exhaust of HCl into the atmosphere. Each launch emits approximately 200 tons (200,000 kg) Cl out of a total fuel load of 1,700 tons (750 tons of liquid fuel and 950 tons of solid fuel). About one-third of the



**Life cycles of CFCs, HCFCs and solid rocket fuel ( $\text{NH}_4\text{ClO}_4$ ).** Chlorine can either be carried into the stratosphere by atmospheric transport of CFCs, HFCs and by Space Shuttle exhaust. The main difference between CFCs and HCFCs is that most (~90%) HCFCs are removed in the troposphere by hydroxyl radicals (OH). In the case of Space Shuttle, approximately one-third of the fuel ( $\text{NH}_4\text{ClO}_4$ ) is burnt in the stratosphere, producing HCl.

solid rocket motor exhaust, or 68 tons of Cl, is deposited in the stratosphere (defined here as altitudes above 15 km) per launch (see figure). The Shuttle emissions are removed from the stratosphere with a residence time of 3 years. The average integrated chlorine loading in the stratosphere per kton  $\text{NH}_4\text{ClO}_4$  used as fuel is given by

$$\text{IC}_{\text{SS}}(t) = 0.318 (1.0 - \exp[-t/3]) \quad (4)$$

p.p.t.v.-year

The behaviour of the chlorine loading is more comparable to that of HCFCs, as the atmospheric recovery is relatively rapid following cessation of emissions.

The time-dependent ODP is approximately  $\text{IC}_{\text{SS}}(t)/\text{IC}_{\text{F-11}}(t)$ . The ratio of the lead coefficients gives the steady-state ODP value of 0.14. Note that the calculated ODP remains above 0.2 for 50 years. As is also obvious from the equations, the time-dependent ODP is very large for a short time after emission because the impact of the Space Shuttle is immediate whereas CFC-11 (and other HCFCs) take years to reach the stratosphere.

One ambiguity in using equation (4) to calculate the ODP for the rocket fuel lies in what is meant by "emission of 1 kton". If we restrict ourselves to the emission of stratospheric HCl (the primary form of chlorine in the exhaust), then the coefficient in equation (4) is larger and the steady state ODP becomes 1.3. On the other hand, if we consider the emissions to be the total fuel load (liquid plus solid),

then the coefficient in equation (4) scales down and the ODP becomes 0.05. There is no precedent for using a similar definition to reduce the ODP of a CFC used as an aerosol propellant (when the ODP is defined as the original ODP discounted by the weight fraction of the CFC in the spray-can product).

How do the total anticipated emissions from the Space Shuttle launches compare with chlorine loading expected from the substitute HCFCs to be used in refrigeration, or against the CFC background expected over the next decades? Previous assessments<sup>6,14</sup> showed that a launch rate of one Shuttle per month over several years would increase stratospheric chlorine loading by about 3 p.p.t.v. (as much as 10 p.p.t.v. at northern mid latitudes). This could be compared with a chlorine loading of about 3,000 p.p.t.v. generated by the CFCs emitted over the past several decades. Because the chlorine loading is expected to decrease, we believe that it is more appropriate to use another measure. A 10 p.p.t.v. contribution is comparable to a loading of 30 p.p.t.v. expected from a continuous annual release of 50 kton of a typical HCFC (1 Cl atom, 10-year lifetime, atomic weight 100), and of the same order as the contribution from the annual use of CFCs and HCFCs by some developing countries.

### Application-sensitive chemicals

Before the establishment of recovery practice for recycling or destruction, the



applications of the CFCs (apart from use as feedstock to produce other chemicals) are such that, except for a time delay of up to a few decades (the so-called banking time), all the material produced is eventually released to the atmosphere (very little is destroyed during applications). The Protocol recognizes this and explicitly exempts the amount used in feedstock, and the amount recaptured for recycling and destruction, from being counted as production in the basket approach within each group. In extending the control strategy to a wider class of compounds, however, we find that some usage may result in having only part of the applied material released to the atmosphere because the compound is partially destroyed in the application. For example, when used as a soil fumigant, about 10 to 50% of the  $\text{CH}_3\text{Br}$  undergoes prompt chemical transformation in the soil and is not released to the atmosphere<sup>8</sup>. These values are based on several model calculations and limited measurements. If a large fraction of the compound were destroyed (without emitting Cl or Br into the atmosphere) during its application and a portion of the chemical emitted to the atmosphere were destroyed by exchange with the Earth's surface or the ocean, should we extend the existing definition of ODP which currently only focuses on the atmospheric portion of a compound's life cycle? By defining the ODP over an extended life cycle, we will be able to include whatever effects are associated with the differences in life cycles and thus provide a more meaningful measure of relative ozone depletion. In any case, the amount destroyed during any application should be quantified and counted as 'amount destroyed' by approved technologies.

### Small-use chemicals

Although the Protocol contains provisions to exempt small medical use of existing CFCs, there is no uniform guideline for assessing new drugs. Many specialized pharmaceuticals contain bromine. Because a Br atom is about 40 times more efficient than a Cl atom in removing ozone<sup>3</sup>, the ODPs of these compounds are about the same as the CFCs even if their lifetimes are many times shorter. Although these compounds do pose a short-term risk to the ozone layer, they are expected to have very small annual production (a few tons). How should production of such unique, beneficial drugs be controlled while protecting the ozone layer?

### Proposed strategy

We have discussed here several ideas that have been considered implicitly in formulating the control strategy in the Protocol to limit the chlorine loading expected from long-lived halocarbons used in traditional industries. As the list of potential

ozone-depleting substances grows, the protection of the ozone layer may require extension of the Protocol to the release of Cl and Br in the stratosphere in connection with other human activities. To this end, it is necessary to follow an approach that relies on our scientific understanding of stratospheric ozone and the life cycles of the compound in question. The aim is to provide sound, long-term protection for the ozone layer, while having a less capricious impact on technological development. The guidelines and the rationale follow.

■ Register all industrially produced chlorine- and bromine-bearing compounds that can potentially put reactive chlorine and bromine species into the stratosphere, and determine their ODPs taking into account the product life cycles. We need to identify all potential sources of stratospheric chlorine and bromine.

■ Ban or place stringent limits on the production of those compounds with long atmospheric lifetimes (lifetime threshold to be determined). Using only short-lived compounds will allow for a reasonably rapid recovery of the atmosphere should unforeseen factors, such as volcanoes or other unanticipated global changes, suddenly enhance chlorine-driven ozone depletion.

■ Set a predetermined goal for the recovery of stratospheric ozone, and thus a schedule for stratospheric chlorine levels (or chlorine plus bromine equivalent measure). Define a market basket approach for the use of all Cl- and Br-bearing chemicals so that the ODP-weighted emissions meet this schedule.

The idea of an ODP-weighted basket is not new. It is part of the protocol in calculating the production within each group of chemicals and in specifying the limit for substitutes. What is different here is the extension to other, non-CFC sources of stratospheric pollution by using an ODP-weighted emission basket that can be tied directly to stratospheric chlorine loading. For the CFCs and HCFCs, the ODP-weighted emission gives results equivalent to those for the ODP-weighted production after exempting feedstock, and recapture for recycling and destruction. However, the particular use of a compound becomes important when evaluating non-volatile chlorinated chemicals, such as the  $\text{NH}_4\text{ClO}_4$  in Space Shuttle fuel, which are not expected to put reactive chlorine into the stratosphere unless they are used in highly specific ways. For example, static tests of the shuttle engine on the ground should not affect stratospheric ozone. In such cases, the distinction between production and emission becomes important. Extending the ex-

isting definition of ODP over a product life cycle could help reconcile this difficulty.

### Future prospects

The broad outline proposed here aims to impose taxes or to place caps on the different end-uses of a compound in which part of the chemical is destroyed. It allows for limited use of any Cl- and Br-bearing compounds considered beneficial to society whose values and costs are to be determined by market forces or by national priorities.

The global limit on ODP-weighted emissions can be agreed by parties to the Protocol and each country can choose the manner in which it limits combined emissions of chloro- and bromocarbons (by taxation, by free-market auction of permits or by explicit caps for specific compounds). The permit trading being used for controlling  $\text{SO}_2$  emissions in the United States<sup>15</sup> may be a viable approach to setting the pollution fees involved in releasing ozone-depleting chemicals. Chlorine-free rocket fuels are being developed as alternatives to  $\text{NH}_4\text{ClO}_4$  (ref. 16). The pollution fee system could help to translate the environmental pay-off into more concrete terms.

The rationale for the control of other ozone-depleting substances unrelated to the chlorine and bromine chemical cycles (for example, nitrogen fertilizer, which produces nitrous oxide as a degradation product) is far more complicated and, we believe, needs to be understood further before being regulated. □

*Malcolm K. W. Ko and Nien-Dak Sze are at Atmospheric and Environmental Research Inc., 840 Memorial Drive, Cambridge, Massachusetts 02139; Michael J. Prather is in the Department of Earth System Science, University of California, Irvine, California 92717, USA.*

1. Wuebbles, D. J. *J. geophys. Res.* **88**, 1433–1443 (1983).
2. Fisher, D. A. *et al. Nature* **344**, 508–512 (1990).
3. Solomon, S. *et al. J. geophys. Res.* **97**, 825–842 (1992).
4. Solomon, S. & Albritton, D. L. *Nature* **357**, 33–37 (1992).
5. *Scientific Assessment of Ozone Depletion: 1991* Ch. 6 Global Ozone Research and Monitoring Project Rep. No. 25 (World Meteorological Organization, Geneva, 1992).
6. *Scientific Assessment of Ozone Depletion: 1991* Ch. 10 Global Ozone Research and Monitoring Project Rep. No. 25 (World Meteorological Organization, Geneva, 1992).
7. *Scientific Assessment of Ozone Depletion: 1991* Ch. 9 Global Ozone Research and Monitoring Project Rep. No. 25 (World Meteorological Organization, Geneva, 1992).
8. *Methyl Bromide: Its Atmospheric Science, Technology and Economics* Montreal Protocol Assessment Supplement (United Nations Environment Programme, Nairobi, 1992).
9. Prather, M. J. & Watson, R. T. *Nature* **344**, 729–734 (1990).
10. *Scientific Assessment of Ozone Depletion: 1991* Ch. 8 Global Ozone Research and Monitoring Project Rep. No. 25 (World Meteorological Organization, Geneva, 1992).
11. Holton, J. R. *J. atmos. Sci.* **47**, 392–395 (1990).
12. Ko, M. K. W. *et al. J. geophys. Res.* **96**, 7547–7552 (1991).
13. *Chem. Engng News* p. 14 (16 August 1993).
14. Prather, M. J. *et al. J. geophys. Res.* **95**, 18583–18590 (1990).
15. *Science* **260**, 1884–1885 (1993).
16. *Chem. Engng News* p. 18 (17 January 1994).

# Time machines still over horizon

**A sixty-year old calculation by Enrico Fermi is discovered to be in error, and inter-atomic signalling between atoms to be potentially faster than light. But this is not a sign that time machines that defy causality can now be built.**

THE idea that the laws of physics are positive interdictions of certain desirable phenomena is much resented. Thus the notion that gravitational forces act indiscriminately, in the direction of the local gravitational field, in proportion only to the mass involved is offensive to those who would take thought and levitate. The common denial that machines capable of doing work (in the technical sense of producing net energy) cannot do so perpetually is often regarded as an intolerable abridgment of the freedom of the human spirit. Not to be able to measure the position and momentum (or any other pair of conjugate variables) simultaneously, as the Uncertainty Principle would have it, is similarly regarded as an affront to the dignity of the human species, even a lie in which malevolent researchers have conspired.

Dr Gerhard C. Hegerfeldt from the University of Göttingen may therefore unwittingly have given a hostage to fortune with his reexamination of the treatment of a problem first tackled in 1932 by Enrico Fermi, at the suggestion of Heisenberg, and now discovered to be flawed (*Phys. Rev. Letts* **72**, 596–599; 31 January 1994). For Fermi had set out to show that the influence of an excited atom on an (unexcited) neighbour would be communicated no more quickly than the speed of light, and had satisfied the referees of *Reviews of Modern Physics* (**4**, 87; 1932) that his calculations were correct. If, now, Fermi is discovered to have been mistaken, and the influence is communicated superluminally, then time machines must surely be just around the corner.

In reality, the case of Fermi's excited atom is neither the scandal nor the opportunity it appears, but an instructive illustration of how much has been learned since 1932. The problem is simply stated. Let there be an atom  $A$  excited into its first electronic state and another  $B$  at a distance  $R$ ; further suppose that the space between the two is uncomplicated by the presence of radiation. The question is to know the probability that  $B$  will be excited by the photon produced by the decay of  $A$  from its excited state to the ground state, and in particular to know the least time-delay. Fermi's conclusion was that the time-lag would be at least  $R/c$ , where  $c$  is the velocity of light.

In 1932, the obvious way to calculate the problem was to use Schrödinger's time-dependent wave equation to calculate the effect on  $B$  of the pulse of electromagnetic energy given off when  $A$  decays. The complication of the problem is that the pulse is not instantaneous, but spread out over a

(short) length of time related to the line-width of the corresponding spectral line by the Uncertainty Principle. That means that the pulse of energy sensed by  $B$  is not that of a single frequency, but is distributed over all possible frequencies. What seems to have gone wrong in Fermi's treatment is an approximation used to represent a mathematical integral over all positive frequencies. Lesser mortals than Fermi should not take too much pleasure in that.

That something was wrong with Fermi's treatment of this seemingly simple problem appears to have been spotted in the late 1940s. Hegerfeldt gives M. I. Shirokov the credit for telling (in 1962) precisely what the error was, although there have apparently been attempts to save the appearances even in the past few years, presumably incorrectly. Hegerfeldt now reckons to have provided a correct treatment which, among other things, shows that there is a finite if small probability that  $B$  will "notice" the decay of  $A$  long before the interval  $R/c$  has elapsed. Does that not imply that causality in the sense of Einstein's relativity has collapsed, and that the time-machine builders should be licensed to get on with their important work?

That, luckily, is a premature conclusion. Although the formal part of Hegerfeldt's paper is an elegant argument about the representations of operators in an appropriate Hilbert space, which demonstrates yet again that it is possible by elegant mathematics to extract something substantial from nothing, the general interest of his paper is his reformulation of the question Fermi asked himself in 1932. First, the atoms  $A$  and  $B$  are taken to be strictly independent systems, each with its own system of energy levels unaffected by the other. Second, the intervening vacuum is supposed to be free from photons that may spuriously excite  $B$ , whether or not  $A$  has decayed. But each of these assumptions is, at some level, false. Fermi, in short, had tackled an unphysical problem.

Take the business of the separateness of the two atoms. It is of the essence of quantum mechanics that the electrons in an atom are infinitely spread out physically. The probability of finding one of  $A$ 's electrons at some distance may be a rapidly decreasing function of the distance, but it is not strictly zero short of infinity. The assumption that the two atoms are strictly independent is thus strictly false, even though it may be a good starting point from which to embark on a more refined calculation. Accordingly,

it should be no surprise that  $B$  begins to notice the decay of  $A$  before light could have travelled the intervening distance. Of course, the magnitude of the probability that  $B$  will respond superluminally does decrease rapidly with the separation distance, but that is not the hole in causality that the time-machine fraternity would concentrate upon.

The assumption that the intervening vacuum is free from photons is even less plausible these days. For one thing, the treatment of the mathematical infinities that arise in the calculation of the properties of electrons in quantum electrodynamics (whimsically, "QED") is crudely the equivalent of surrounding all electrons with a kind of photon cloud, substantially contributing to the overlap between one atom and even a relatively distant neighbour. The "polarization of the vacuum" is the name for this phenomenon, for which Feynman, Schwinger and Tomonaga, with the help of apologists such as Dyson, were responsible late in the 1940s.

QED creates further complications, not surprising when people are forever unearthing evidence that the properties of the electronic states of atoms are affected simply when they are placed in cavities capable of sustaining only some modes and frequencies of electromagnetic radiation. Hegerfeldt puts the philosophical point that, in these circumstances, it may make no sense to say that  $B$  can be put into its theoretical ground state, calculated for a strictly isolated atom, when  $A$  is somewhere in the neighbourhood, exerting an influence through the intervening vacuum.

Hegerfeldt acknowledges that all three arguments are simply different ways of saying that the two atoms are not strictly separate systems. What his own argument, formal and thus general, does not provide is an estimate of the magnitude of the probability of superluminal signalling between two supposedly independent systems of this kind. It will be interesting to see, when that algebra has been done, whether the superluminal signal is a more rapidly decreasing function of the distance than is the causal signal after an interval of  $R/c$ . That would be the general expectation.

Whatever the truth, the serious moral in Fermi's error is that approximations used as the basis for calculations, however convenient, may essentially omit some element of the reality they are meant to represent. Meanwhile, those planning to build time machines will have to wait a little longer for the details to emerge.

John Maddox



# Looking for galaxies in all the wrong places

John J. Salzer

ASTRONOMERS have been faced with a serious problem for the past several years: the numbers of galaxies detected at very faint flux levels are too numerous by a factor of five or more to be explained by extrapolating the observed density of galaxies seen in the nearby Universe to larger distances. Explanations for this excess have been offered, but usually require either abandoning the standard view of how galaxies form and evolve, or modifying cosmology in a fundamental way. McGaugh<sup>1</sup>, on page 538 of this issue, offers an alternative explanation: the problem may not be that there is an excess of faint, distant galaxies, but rather that too few galaxies are being catalogued locally. The missing galaxies are those with a very low surface mass density of stars.

Attempts were made to use galaxy counts as cosmological probes from the very earliest days of extragalactic research<sup>2</sup>. It was soon realized, however, that the evolution of the stellar populations that make up the galaxies could change their properties (luminosity, colour) considerably in the billions of years it takes the light to reach us from the most distant objects. Modern work on galaxy counts makes use of sensitive detectors which have pushed surveys to unprecedented depths. But the interpretation of these counts relies on the use of sophisticated models, with many adjustable (and usually poorly known) parameters, which attempt to account for both the effects of galaxy evolution and those of the adopted cosmological model on the observed counts.

## Excess

Recent surveys<sup>3-5</sup> reveal large excesses of faint galaxies above the numbers predicted from the observed density of nearby galaxies, a standard cosmological model, and the assumption that galaxies haven't changed drastically over the past several billion years. This excess galaxy population has colours that are much bluer than normal, indicating a young mean stellar age. Furthermore, redshift surveys<sup>4,6,7</sup> show that these objects lie at intermediate distances and have fairly modest luminosities. The failure of the so-called 'no evolution' models to explain the observed number counts has led to many and varied attempts to reconcile the situation. As reviewed by McGaugh, all of these 'exotic' solutions have serious flaws. Rather than requiring significant, and in many cases untenable, modifications to our view of galaxy evolution or cos-

mology, he suggests that we have simply missed counting a large fraction of nearby galaxies.

McGaugh's choice for this missing population is low-surface-brightness galaxies. These are galaxies where the luminosity per unit area is substantially below that of the bright spiral and elliptical galaxies which dominate the pages of most galaxy atlases<sup>8</sup>. Until very recently, such objects were virtually impossible to study (or even discover) because their surface brightnesses were fainter than the natural emissions of the dark night sky. But deep, wide-field imaging surveys and follow-up observations are now showing that a large population of galaxies has been hidden from our view<sup>9-12</sup>. Furthermore, these faint objects have blue colours and intermediate luminosities consistent with the excess population of faint galaxies, and are present in sufficient numbers in the nearby Universe that they may actually be more common than their more readily detected high-surface-brightness counterparts.

How could this huge population of galaxies exist and yet remain largely undiscovered until now? The answer has to do with the way galaxies are selected for inclusion in catalogues. Because galaxies are resolved, their light is spread over a finite area of the sky. For a given total flux, a larger galaxy will be less conspicuous than a more compact one. If the light is too spread out, the low-surface-brightness galaxy is lost in the noise produced by the night sky background. McGaugh shows how the selection biases present in nearly all existing bright galaxy catalogues lead to the underrepresentation of low-surface-brightness galaxies in these supposedly complete samples. The bias works in two ways: first, many galaxies are missed from the catalogue entirely, and second, those low-surface-brightness galaxies included in the lists have their total fluxes severely underestimated. On the other hand, the ultra-deep surveys used for faint galaxy number counts are sensitive enough to detect even very low-surface-brightness objects and provide accurate total flux measurements.

The identification of low-surface-brightness galaxies with the excess in the number counts is a much more elegant and simple solution to the faint blue galaxy problem than those previously suggested. It does not require any unusual form of galaxy evolution or non-standard cosmological models to explain the observations. Score one for Mr Occam and his razor.

Furthermore, Koo and collaborators<sup>13,14</sup> have found that models with little or no evolution can in fact explain the observed number counts if they include blue intermediate- and low-luminosity galaxies whose properties are not that dissimilar from the low-surface-brightness galaxies. The problem is that the true extent of the local population of low-surface-brightness galaxies is still largely unconstrained. By their very nature, these objects are difficult to detect and study. Like the rest of society, astronomers have a hard time believing in things that they can't see.

## Evidence

There is ample precedent for the existence of a large population of hidden low-surface-brightness galaxies in the local Universe. Tyson and Scalo<sup>15</sup> used existing dwarf galaxy catalogues and models of the selection effects present in them to predict the existence of a substantial dwarf galaxy population which was virtually unknown. Furthermore, surveys for low-surface-brightness galaxies<sup>9,10</sup> find a high surface density of such galaxies on the sky, and these galaxies represent essentially the full range of galaxy sizes and luminosities: low-surface-brightness galaxies are not all dwarfs.

Indirect evidence for this hidden population comes from the observed space density of intermediate- and low-luminosity star-forming galaxies<sup>16</sup> which have high surface brightness. These objects, which are relatively easy to detect and study, comprise roughly 5 per cent of the known galaxy population outside clusters. However, stellar evolution timescales suggest that only 0.1 to 1 per cent of galaxies with these luminosities should be in the star-forming phase. The most natural interpretation for this discrepancy is that the progenitors of the star-forming galaxies are mainly uncatalogued low-surface-brightness objects which are present in great numbers. After the end of the

1. McGaugh, S. S. *Nature* **367**, 538-541 (1994).
2. Hubble, E. & Tolman, R. C. *Astrophys. J.* **82**, 302-337 (1935).
3. Tyson, J. A. *Astr. J.* **96**, 1-23 (1988).
4. Lilly, S. J., Cowie, L. L. & Gardner, J. P. *Astrophys. J.* **369**, 79-105 (1991).
5. Metcalfe, N., Shanks, T., Fong, R. & Jones, L. R. *Mon. Not. R. astr. Soc.* **249**, 498-522 (1991).
6. Colless, M., Ellis, R. S., Broadhurst, T. J., Taylor, K. & Peterson, B. A. *Mon. Not. R. astr. Soc.* **261**, 19-38 (1993).
7. Cowie, L. L., Songaila, A. & Hu, E. M. *Nature* **354**, 460-461 (1991).
8. Sandage, A. *The Hubble Atlas of Galaxies* (Carnegie, Washington DC, 1961).
9. Impey, C., Bothun, G. & Malin, D. *Astrophys. J.* **330**, 634-652 (1988).
10. Schombert, J. M., Bothun, G. D., Schneider, S. E. & McGaugh, S. S. *Astr. J.* **103**, 1107-1133 (1992).
11. McGaugh, S. S. & Bothun, G. D. *Astr. J.* (in the press).
12. Knezek, P. thesis, Univ. of Massachusetts (1992).
13. Koo, D. C. & Kron, R. G. *A. Rev. astr. Astrophys.* **30**, 613-652 (1992).
14. Koo, D. C., Gronwall, C. & Bruzual, A. G. *Astrophys. J.* **415**, L21-L24 (1993).
15. Tyson, N. D. & Scalo, J. M. *Astrophys. J.* **329**, 618-628 (1988).
16. Salzer, J. J. *Astrophys. J.* **347**, 152-170 (1989).

star-formation burst, the hot, luminous stars evolve away on very short time-scales, and these objects most probably fade back into low-surface-brightness galaxies like those envisaged by McGaugh.

The reader should keep in mind that the suggested identification of the excess faint blue galaxies with low-surface-brightness galaxies is just that: a suggestion. McGaugh has shown that the low-surface-brightness galaxies could account for the excess faint galaxy population. Although his solution to this problem is not the only

one, it's certainly one of the most interesting. Perhaps more important for astronomy is the picture this work paints of the overall galaxy population. Do we really have a complete census of galaxies in the local Universe? Evidence is accumulating for the existence of vast numbers of previously uncatalogued galaxies. This newly uncovered population will require us to think about galaxies in a different light. □

*John J. Salzer is in the Astronomy Department, Wesleyan University, Middletown, Connecticut 06459, USA.*

## ARCHAEOZOOLOGY

# Time for a change

*Paul G. Bahn*

A CENTRAL event in human history in South America was the domestication of the native camelids. But when and how did it take place? Carmen Reigadas of the Instituto de Ciencias Antropológicas in Buenos Aires reports<sup>1,2</sup> a fresh way of tackling these questions. She has developed a technique of examining fleece remains and products which will be of great value in complementing traditional work with bones, as well as research, now in train, involving DNA analysis.

Archaeologists have had great difficulty in distinguishing remains of the domestic camelids (llamas and alpacas) from those of wild forms (guanacos and vicuñas). In consequence, the transition from hunting to herding — that is, from the exploitation of wild camelids to that of domestic forms — is still poorly understood<sup>3</sup>. The ancestor of the two domestic forms is also largely a matter for conjecture, though it was probably not the vicuña, the smallest form with the world's finest wool.

Since the 1970s, hundreds of thousands of camelid bones have been examined, but few, if any, post-cranial skeletal features have been identified that can be used to differentiate between the species. In any case, the bones from many sites are highly fragmented, damaged or weathered. The main criterion of differentiation has been the shape of incisor teeth. But even here there is considerable overlap — the incisors of modern guanacos and llamas are indistinguishable, and alpaca incisors are intermediate in form between those of guanaco/llamas and those of vicuñas.

A different approach was taken by Jane Wheeler<sup>4</sup>. In her study of the abundant camelid bones from the high-altitude

(4,420 m) prehistoric site of Telarmachay, Peru, she found a marked increase through time in the proportion of fetal or neonatal remains. In guanaco and vicuña populations today, 35–40% is normal, but at Telarmachay the figure rose from 45% to 68% and even 73%. The very high mortality rate was ascribed to entero-

## IMAGE UNAVAILABLE FOR COPYRIGHT REASONS

Domestic types — llamas being herded in Bolivia.

toxaemia, a diarrhoea caused by bacteria related to dirty, muddy corrals, which is unknown in wild species and was therefore thought to be a by-product of domestication. On that basis, the layer with 45% was seen as transitional between hunting and herding (between about 6,000 and 5,500 years before present (yr BP)), with herding of fully domesticated forms after about 5,500 yr BP. However, the exact identification of the species involved was not possible.

Carmen Reigadas has opted for a different tack again. One key advantage of many sites, particularly those at high altitude, is their extreme aridity which has permitted the survival of normally perish-

able items — in particular cordage, textiles and fleece — in large quantities. Reigadas has therefore undertaken a microscopic analysis of the fibres and follicles present in these items.

In a study akin to the pioneering work carried out by Michael Ryder on ancient sheep remains in the Old World<sup>5</sup>, she has looked at changes in the fibres through time. Most mammals have two coats: one of short, thin fibres (the under-coat or under-wool), and one of coarser fibres (the kemps, or hair). It has been established that among the consequences of domestication are a decrease in the size of the primary follicles (from which the kemps grow), an increase in the number of the secondary follicles (which house the under-coat), a thinning of kemps and, overall, a less marked double coat. Ryder found that fibre width altered upon domestication, and that a third type of fibre, intermediate between the under-coat and the kemps, appeared.

Reigadas began by taking fibre samples from different parts of the body of the four modern camelids, and measuring each hair's overall diameter and that of its medulla (core). It turned out that the vicuña has a characteristic 'binary' coat:

82% of its fibres are less than 25 µm wide, while the rest are very thick hairs of 66–76 µm. The guanaco also has a binary distribution, but with different measurements: it has fewer thin fibres (70% 26–46 µm), while 20% are very thick. For the llama, as expected, there is a predominance of intermediate fibres (82% 16–46 µm) and a clear lack of either very thick or very thin fibres. It is therefore very different from the vicuña, but has a certain overlap in size with the guanaco: for the vicuña and llama, any fibre under 31 µm is under-coat, whereas the guanaco can have hairs under 31 µm. Conversely, hairs of the llama that exceed 31 µm can overlap with the guanaco's under-coat. Reiga-

das therefore settled on the following upper limits: 31 µm for thin fibres, and 66 µm for intermediates, anything over 66 µm being counted as thick fibres.

As her first case study, she examined finds from the cave of Huachichocana III, excavated in the 1970s by Alicia Fernández Distel. The cave is located at 3,400 m in Argentina's northern Puna; it was periodically occupied over a long time span, and contains faunal material.

Layer E3, with radiocarbon dates from 10,200 to 8,670 yr BP, in the Early Pre-ceramic period, yielded fleece and cordage samples. The fibre measurements corresponded entirely to guanaco (two samples) and vicuña (one sample). There

NHPA/Trevor McDonald



were no intermediate fibres at all. Layer E2, dated to 3,400 yr BP (Late Pre-ceramic), again gave measurements with a characteristic bimodal distribution and no intermediates. There were values for the guanaco, and medulla measurements that pointed to the vicuña. However, layer E1, dated to 1,420 yr BP (the Formative period), contained cordage of which one sample was clearly vicuña, but three others had a predominance of intermediate fibres (46–56 µm), with no very thick hairs, and they clearly came from llamas: their homogeneity points to a characteristic selection of fibres for cordage.

Later layers (D, C) likewise contain a mixture of wild and domestic types, but with domestic forms (intermediate fibres) becoming more dominant, and with increasing homogeneity and more indication of careful fibre selection. Hunting was being complemented and largely replaced by herding, it seems, as shown by a guano layer in D which indicates that the cave was used as a corral.

The fibre results therefore suggest that E1 was a time of transition from hunting to herding, and that in this area domestication was under way well before 1,420 yr BP. This fits well with data from the skeletal remains in the site. Hugo Yacobaccio and Celina Madero<sup>6</sup> endeavoured to identify camelid species by measurement of the width of the proximal epiphysis of the first phalange, and by other osteometric measurements which might reveal overall size (llamas are biggest, vicuñas smallest, the others intermediate). The remains were almost entirely fragmentary, but layer E2 contained one complete skull of an adult. Its incisors were spatulate, indicating either a guanaco or a llama; and an allometric calculation based on jaw measurements gave an estimate of 127 kg live weight. Guanacos weigh only 70–90 kg, so this was clearly a llama (which weigh 90–140 kg).

Hence, the appearance of a domestic form around 3,400 yr BP—a figure which agrees with data from other sites in the region—also confirms the predominance of llama fibres in the next occupation of the site. More broadly, the work shows that fibre analysis will be an invaluable aid at sites where bone remains are absent or too fragmentary to be of use. □

Paul G. Bahn, 428 Anlaby Road, Hull HU3 6QP, UK, is a freelance writer on archaeology.

## Properly preparing promoters

Ken van Holde

How can genes that are normally inactive in a particular cell type suddenly become expressible in response to an external stimulus? There is plenty of evidence that in many cases chromatin structure pre-existing over the gene and its flanking sequences must be disrupted. The critical step may be freeing the promoter region from protection by nucleosomes to allow binding of transcription factors. This kind of competition between factors and nucleosomes has been referred to by Felsenfeld<sup>1</sup> as 'dynamic competition', to distinguish it from the 'pre-emptive competition' which can occur at the time of DNA replication. It seems reasonable that special proteins, distinguished by their ability to rearrange chromatin, might be required. Indeed, as early as 1984 it was shown that an unidentified protein factor could confer hypersensitivity to nucleases on a specific region in a chicken globin promoter<sup>2</sup>. Since then a number of such proteins have been identified<sup>1,3</sup>, but their mechanism of action remains obscure.

In the paper that appears on page 525 of this issue<sup>4</sup>, Tsukiyama, Becker and Wu provide some insights into these matters. The study describes the *in vitro* establishment of nuclease-hypersensitive sites in the promoter of heat-shock gene 70 of *Drosophila*. The example is a curiously appropriate one, for it is in the same promoter that Wu first precisely mapped a hypersensitive site<sup>5</sup>.

Activation of the heat-shock genes is a multi-step process. In *Drosophila* cells the system exists in a 'poised' state, with RNA polymerase II already in place and engaged. Release of the polymerase and start of transcription requires the binding of heat-shock factor (HSF) proteins, whose binding is in turn conditional on binding of the GAGA factor<sup>6</sup>. This protein<sup>7</sup> is a multi-purpose transcriptional activator which binds to (GA/CT)<sub>n</sub> sites in a host of promoters, and genetic studies<sup>8</sup> have shown that the function of GAGA, at least at the heat-shock loci, is probably one of promoting chromatin rearrangement.

But what rearrangement, and how? This is where the latest work<sup>4</sup> comes in. Wu and his collaborators had the advantage of a powerful new *in vitro* system for chromatin assembly, using a *Drosophila* embryo extract<sup>9</sup>. They had already shown that this system, when augmented with an ATP energy source, would produce a regular nucleosome array that blocked transcription from an *hsp70* promoter. The gene for the GAGA factor has recently been cloned<sup>10</sup>, and its expression in a baculovirus system provided protein for activity assays.

In brief, Tsukiyama *et al.*<sup>4</sup> have combined these advances to investigate the effect of GAGA factor on the *hsp70* locus in reconstituted chromatin. In the absence of GAGA factor, the assembly system places a nucleosome (not necessarily well positioned) overlapping two of the four GA/CT sequences in the promoter. The addition of partially purified GAGA factor, before, during or after completion of chromatin assembly, led to the formation of a nuclease-hypersensitive site in this locus, and seemed to help position nucleosomes in the surrounding regions. The requirement for GAGA factor was confirmed by the use of specific antibodies. The addition of histone H1 (absent in the *Drosophila* embryo extract) inhibited but did not completely eliminate nucleosome disruption. These observations are wholly consistent with the prediction by Lu *et al.*<sup>8</sup>, from genetic analysis, that "GAGA factor and HSF appear to have distinct roles in gene regulation, the former setting up chromatin structure and the latter exploiting it".

A most intriguing aspect of the new study is the observation that ATP greatly facilitates the modifications in chromatin structure, and may even be necessary for them to occur. If chromatin was assembled, and the ATP-hydrolysing enzyme apyrase then added, GAGA factor became ineffective; depletion of ATP had the same effect. Conversely, addition of ATP to depleted systems restored the effectiveness of GAGA factor, but addition of non-hydrolysable ATP analogues did not.

This is the first time that an exogenous energy source has been shown to be involved in chromatin disruption. Just how and where the energy is exerted remains a mystery. GAGA factor does not appear to bind ATP, nor to possess ATPase activity. Presumably some as yet unknown component in the embryo extract is involved, and identifying it should throw up some exciting possibilities. It

1. Reigadas, M. C. *Arqueología* (Buenos Aires) **2**, 9 (1993).
2. Reigadas, M. C. *Zooarqueología de Camélidos* (in the press).
3. Kent, J. thesis, Washington Univ., St. Louis (1982).
4. Wheeler, J. C. in *Animals and Archaeology* Vol. 2, *Early Herders and their Flocks* (eds Clutton-Brock, J. & Grigson, C.) 395–410 (BAR Int. Series 202, Oxford, 1984).
5. Ryder, M. in *The Domestication and Exploitation of Plants and Animals* (eds Ucko, P. J. & Dimbleby, G. W.) 495–521 (Duckworth, London, 1969).
6. Yacobaccio, H. D. & Madero, C. M. *Arqueología* (Buenos Aires) **2**, 149–188 (1993).

1. Felsenfeld, G. *Nature* **355**, 219–224 (1992).
2. Emerson, B. M. & Felsenfeld, G. *Proc. natn. Acad. Sci. U.S.A.* **81**, 95–99 (1984).
3. Axelrod, J. D., Reagan, M. S. & Majors, J. *Genes Dev.* **7**, 857–869 (1993).
4. Tsukiyama, T., Becker, P. B. & Wu, C. *Nature* **367**, 525–532 (1994).
5. Wu, C. *Nature* **286**, 854–860 (1980).
6. Lis, J. & Wu, C. *Cell* **74**, 1–4 (1993).
7. Biggin, M. D. & Tjian, R. *Cell* **53**, 699–711 (1988).
8. Lu, Q., Wallrath, L. L., Granok, H. & Elgin, S. C. R. *Molec. cell. Biol.* **13**, 2802–2814 (1993).
9. Becker, P. B. & Wu, C. *Molec. cell. Biol.* **12**, 2241–2249 (1992).
10. Soeller, W. C., Oh, C. E. & Kornberg, T. B. *Molec. cell. Biol.* **13**, 7961–7970 (1993).
11. Tremethick, D. J. & Frommer, M. *J. Biol. Chem.* **267**, 15041–15048 (1992).
12. Tremethick, D. J. & Drew, H. R. *J. Biol. Chem.* **268**, 11389–11393 (1993).

should be noted that a protein factor that confers ATP-dependent spacing has been identified in the *Xenopus* chromatin assembly system<sup>11</sup>, and that phosphorylated HMG14 or 17 can fulfil part of this function<sup>12</sup>. Perhaps a similar factor (or factors) exists as a part of the *Drosophila* system, and is able, upon phosphorylation, to support rearrangement of nucleosomes. It will also be of greatest importance to see whether or not GAGA factor plus heat-shock factor can lead to the restoration, *in vitro*, of the transcriptional capability of the *hsp70* promoter.

Potentially important as these results may be, it must be remembered that they come from experiments *in vitro*. In the fly, GAGA appears to be present on the heat-shock promoter even in the uninduced state, rendering the *in vivo* promoter hypersensitive to nuclease. So the kind of changes described here must relate to the conversion of an uninducible

promoter to an inducible one at some earlier stage of development.

Perhaps a major implication of both this kind of study and studies of chromatin assembly systems is the existence of a previously undefined functional class of proteins — proteins whose role is to promote rearrangements in chromatin structure. Such proteins would be expected to be of moderate abundance and to bind to many sites in chromatin. Tsukiyama, Becker and Wu<sup>4</sup> provide hybridization evidence that GAGA factor meets at least the second requirement. To distinguish these factors from the transcription factors that interact directly with the polymerization machinery, one might call them 'chromatin modifying factors'.

Ken van Holde is in the Department of Biochemistry and Biophysics, Oregon State University, Corvallis, Oregon 97331, USA.

## QUANTUM COMPUTING

# Shannon's theorem revisited

Artur Ekert

In the past few years there has been an explosion of theoretical and experimental innovations which, their discoverers claim, are creating a fundamental new discipline: a distinctively quantum theory of information. Quantum physics allows, for example, the construction of qualitatively new types of logic gates, absolutely secure cryptosystems and the possibility of cramming two bits of information into a single physical bit, to mention only the most spectacular achievements<sup>1-4</sup>. The most celebrated theorems on which the whole edifice of information theory was built must be modified to suit the rules of quantum data processing. New work by Benjamin Schumacher shows that even the Shannon noiseless coding theorem, regarded as a basic precept of the field, must be reformulated to accommodate the qualitatively new quantum coding methods<sup>5</sup>.

## Classical coding

The Shannon noiseless coding theorem answers one of the fundamental questions in communication theory: what is the ultimate limit to data compression? The basic idea behind the theorem is simple: data compression can be achieved by assigning a short description to the most frequent outcomes of the data source and a necessarily longer description to the less frequent outcomes<sup>6</sup>.

To take an example, suppose you wanted to rewrite all Shakespeare's comedies using, instead of the normal English alphabet, the binary alphabet composed only of the two characters 0 and 1. Your

first task would be to find a code to convert all the characters appearing in the English text to their respective strings of zeros and ones. Of course, you would like to keep the final huge string of bits as short as possible, so you have to think carefully before committing yourself to a specific code. The frequency of letters in English is far from uniform. The most common letter, E, has a frequency of about 13 per cent, whereas the least common letters, Q and Z, occur with a frequency of about 0.1 per cent. It is reasonable then to assign the shortest string to E and the longest strings to Q and Z. After playing with various codes you would soon realize that it is even better to encode whole blocks of letters together rather than single letters separately, and your final choice would probably be to use block codes with an average number of bits per letter of the order of 2.8. Naturally, you would like to know whether you can do better than that, perhaps by trying more sophisticated binary codes. Shannon's theorem gives the lower bound on the compression.

In a more precise formulation, Shannon's noiseless coding theorem asserts that, given a data source which emits symbols (letters)  $S_1, S_2, S_3, \dots, S_n$  with probabilities  $p_1, p_2, p_3, \dots, p_n$  respectively, each symbol emitted being chosen independently of all other symbols, there exists a binary code which gives the best data compression. The average number of bits per symbol cannot be less than the 'Shannon entropy' of the source, given by  $H = -p_1 \log_2 p_1 - p_2 \log_2 p_2 - p_3 \log_2 p_3 \dots - p_n \log_2 p_n$ .

Knowing this formula and the probabilities, you could in principle estimate the lower bound of the binary code. In practice, however, you may come across many problems. For example, symbols generated by the source may depend on each other, probabilities of occurrence of various symbols can be only estimated with some finite precision, and so forth. Nevertheless, even the rough approximation of the compression factor may be useful. This way you can convince yourself that at least to a first-order approximation you cannot compress Shakespeare's comedies more densely than 2.8 bits per letter. This analysis applies, of course, to any other English text (unless you have to encode something bizarre such as the 267-page novel *Gadsby*, by Ernest Vincent Wright, in which the author deliberately makes no use of the letter E).

## Quantum coding

From the physical point of view, the Shannon theorem allows one to estimate the minimal number of physical resources — two-state systems or physical bits — necessary for a faithful representation of a sequence of symbols. Classical communications theory could rely on the Shannon theorem simply because no quantum effects were directly employed for coding, signalling, transmission or detection. Why should quantum physics make a difference?

Shannon's noiseless coding theorem implicitly assumes that we can always reliably distinguish between different symbols. Different symbols could mean different quantum states of a specific physical carrier of information. Nature, however, follows the quantum rules. These allow for two different, perfectly distinguishable symbol-state preparations to produce two different symbol-states that cannot be reliably distinguished.

Quantum states are described mathematically as vectors in a certain vector space called Hilbert space. If a given physical object can be put into  $N$  perfectly distinguishable states then any state of this physical object can be described as a vector in  $N$ -dimensional Hilbert space. The rules of quantum mechanics say that only states represented by mutually perpendicular vectors can be perfectly distinguished from each other. For non-perpendicular symbol-states, any detection procedure is imperfect and gives either wrong identifications or inconclusive results.

Suppose now that we want to encode messages from a quantum source which produces particles in quantum states  $S_1, S_2, S_3, \dots, S_n$  with probabilities  $p_1, p_2, p_3, \dots, p_n$  respectively, each quantum state emitted being chosen independently of all other states. If the signal states  $S_i$  are mutually perpendicular the classical data compression techniques can be directly



applied. But in general the states are not orthogonal to each other, so even though we know the list of possible states and their *a priori* probabilities, there is no reliable way of identifying the signals with certainty. What is the minimum number of resources — the physical bits — for the coding in this case? Here the physical bit is any quantum object for which the state is described by a vector in two-dimensional Hilbert space. Quantum physical bits are also called 'qubits', so we are trying to find the minimum number of qubits per quantum symbol.

Schumacher's new work<sup>5</sup> has proved that there is a lower bound of number of qubits per symbol and this bound can be less than Shannon's entropy. The quantity that determines the bound is known as von Neumann entropy and depends both on the *a priori* probabilities of the symbols and on the vectors representing the symbol states. The von Neumann entropy is smaller than the Shannon entropy except when the symbol states are mutually orthogonal, in which case the two entropies are equal.

One way to think about the quantum generalization of the Shannon theorem is to consider the coding of longer and longer blocks of symbols. If each symbol-state is described by a vector in the  $N$ -dimensional Hilbert space then a block of  $k$  state-symbols can be described as one vector, but this time in  $N^k$ -dimensional Hilbert space. Because there are  $n$  symbols, then each block of  $k$  symbols can be in one of the  $n^k$  states, each state being a vector in the  $N^k$ -dimensional Hilbert space. But because of the different probabilities associated with different symbols, some blocks of symbols will have negligible probabilities of occurrence. It turns out that the vectors representing the typical blocks with the non-negligible probabilities lie with pretty good accuracy in the  $2^{Dk}$ -dimensional subspace of the  $N^k$ -dimensional Hilbert space where  $D$  is the von Neumann entropy of the source. That is why, for long blocks, we need on average about  $D$  qubits per symbol.

This generalization of the Shannon theorem comes just in time. The technology is ripening to the stage where it can support quantum data processing. Experimental progress in quantum optics enabled us to construct the first quantum cryptosystems. If quantum computation is to be the next experimental target, quantum coding will be essential. □

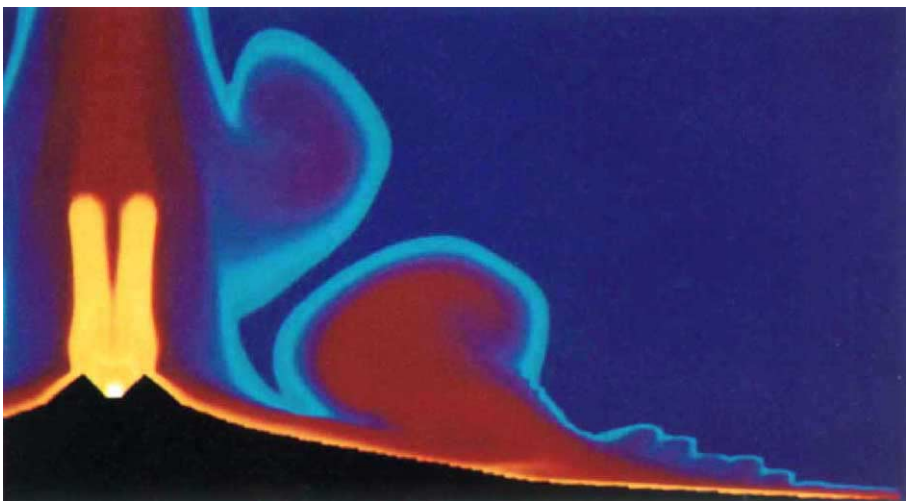
Artur Ekert is at the Clarendon Laboratory, University of Oxford, Parks Road, Oxford OX1 3PU, UK.

## From models to reality

Michael F. Sheridan

In a volcano's eruptive repertoire, some of the most violent phenomena are the giant plinian eruption columns that pierce the tropopause and their associated lateral pulses of hot mixtures of particles and gas. They not only devastate inhabited areas near the volcanoes but may also affect climate worldwide. Understanding the physics of such events and quantifying the movement of the ash clouds associated with giant volcanic eruptions has been the goal of volcanological research during the past quarter of a century.

The most ambitious and sophisticated numerical models of pyroclastic flow eruptions so far are being developed by the Gruppo Nazionale per la Vulcanologia in Pisa. On page 551 of this issue, F. Dobran and colleagues demonstrate the models' abilities: shown here is a snapshot of detailed particle concentrations in a pyroclastic flow moving down the flanks of Vesuvius. This model realistically simulates the generation of an elutriation plume (also known as a phoenix cloud), in which the finer, lighter particles are borne



Raising a dust — plumes of volcanic particles billow upwards, five minutes into this computer simulation of Vesuvius in eruption.

One fruitful line of investigation is to use large computers to make dynamic models of multi-phase systems with the objective of realistically representing natural systems. An early attempt to model large volcanic explosions (K. H. Wohletz *et al.* *J. geophys. Res.* **89**, 8269–8285; 1984) used a two-dimensional axisymmetric numerical simulation based on a shock tube analogy. The most important result was the quantification of parameters related to the initial blast wave produced during the first 60 seconds of giant volcanic explosions. Later models used time-dependent, two-phase compressible Navier–Stokes equations to simulate the first few minutes of continuous volcanic jets (G. A. Valentine and K. H. Wohletz *J. geophys. Res.* **94**, 1867–1887; 1989). These models mimicked the development of plinian eruption columns and clarified conditions that relate to column collapse and determine how material in pyroclastic flows will be distributed. By introducing large obstacles in the finite-element grid, Valentine and co-workers determined the effect of caldera margins on general flow properties (G. A. Valentine *et al.* *Geol. Soc. Am. Bull.* **104**, 154–165; 1992).

upwards by the air currents where the volcano slope flattens.

But how do we use such elegant models to reduce the loss of life near dangerous volcanoes? If Vesuvius were to produce an eruption with pyroclastic flows such as the one pictured here, between 250,000 and 2,500,000 people would be in jeopardy. Evacuation of such numbers from the Naples area, even if they had ample warning, seems improbable. Ideally, future research on this question would bring in teams of policy makers, city planners, engineers, public safety officers and medical experts in addition to geoscientists. One goal of this effort could be to develop low-cost, efficient shelters that would be aerodynamically designed to withstand the heat and force of simulated flows. Hazard planning near volcanoes could then include an adequate number of safe areas, man-made or natural, close enough to the inhabited areas for people to be able to walk to shelter just for the short duration of the most dangerous periods. □

Michael F. Sheridan is in the Department of Geology, State University of New York at Buffalo, New York 14260, USA.

1. Stewart, I. *Nature* **353**, 384 (1991).
2. Ekert, A. *Nature* **358**, 14 (1992).
3. Sudbery, T. *Nature* **362**, 586 (1993).
4. Bennett, C. H. *Nature* **362**, 694 (1993).
5. Schumacher, B. *Phys. Rev. A* (in the press).
6. Welsh, D. *Codes and Cryptography* (Clarendon, Oxford, 1988).

# Mergers make more sense

M. G. Edmunds

NEW space-based and ground-based observations<sup>1,2</sup> may have removed a substantial objection (perhaps the main objection) to the theory that elliptical galaxies formed as a result of the merger of disk-like spiral galaxies. This could at last force out the traditional view that most galaxies form as single objects which separated out from the expanding early Universe, and replace it with the belief — which has been growing for some years — that the history of a galaxy involves the violent combination of several pre-formed or forming sub-systems.

The basic idea that elliptical galaxies could be made from spirals began to gain a following with the theoretical work of Toomre<sup>3</sup> in the 1970s. During a collision, the effect of the rapidly varying gravitational field can cause the stars to share their dynamical energy around, violently relaxing the product galaxy to a more

uniform spheroidal or ellipsoidal shape. Much of the gas in the disks would experience shocks and be persuaded to form stars. Subsequent supernovae and stellar winds would help the escape of most of the remaining gas.

This picture met a real difficulty when van den Bergh<sup>4</sup> perceptively remarked that if it were true, then it is surprising that elliptical galaxies have considerably more globular clusters per unit light output than do spirals. In our own Galaxy, the globular clusters are probably the oldest known objects, beautiful near-spherical clusters of a hundred thousand or even a million old stars, dating back to at least 12–15 billion years ago. In any collision, two spirals might just be able to hang on to most of their globular clusters, but the galaxy produced must have the same or fewer clusters per unit light output than the parent spirals. As ellipticals have, on

average, over twice as many globulars per unit light output as do spirals, this makes it difficult to see how ellipticals can be the children of spirals.

The way out is to follow up<sup>5</sup> the suggestion that new globular clusters are formed in the mergers. At first acquaintance this idea seems almost heretical, given the undisputed great age of our Galaxy's globulars — it is almost like proposing that Egyptian pyramid building has continued right up to the present day. Perhaps we should not be so unprepared, as there are known to be bright young spherical clusters in the nearby Magellanic Cloud's irregular galaxies. These, known as 'populous blue clusters', show many, if not quite all, the characteristics that imply that they will look just like our Galaxy's globular clusters after another 10 billion years of ageing.

Possible young globular clusters have been seen<sup>6</sup> in the galaxy NGC3597, but the ground-based observations did not have sufficient spatial resolution to demonstrate that they were as physically compact as conventional globulars. Another

## OBITUARY

### G. P. S. Occhialini (1907–93)

In a sad coincidence, G. P. S. (Beppo) Occhialini died on 30 December last year within a few weeks of Bruno Rossi and a few months of Bruno Pontecorvo, three of the greatest Italian physicists of the same cultural generation. Occhialini graduated in Florence in 1929, and under the influence of his father Augusto, also a professor of physics, at 24 he joined the Cavendish Laboratory in Cambridge, under the great P. M. S. Blackett.

He brought to the Cavendish the coincidence counter technique ('alla Rossi') and applied it to the Wilson chamber, which until then had been triggered randomly and thus only rarely captured a decent picture of a cosmic-ray track. The new circuit was an immediate success: "One on each, Beppo!" Blackett exclaimed. Next came the famous picture of the first electromagnetic shower and the confirmation of the discovery of the positron by C. Anderson, who narrowly beat them to publication.

After a few years in the increasingly difficult climate of fascist Italy, he went to Brazil to work at São Paulo and later disappeared in the Itatiaya mountains to wait out the Second World War. Just before its end, he emerged to move on to the second great British adventure of his life: Bristol, the Wills Laboratory and C. P. S. Powell. There he immediately grasped the potential of photographic emulsions for elementary particle work. After researching with the photographic company Ilford on how to increase their 'half tone' plate sensitivity, Occhialini personally exposed a group of the new

plates at the Pic du Midi in the Pyrenees. This was during the course of a speleological campaign, another of his great passions. When the plates were developed, in Powell's words, "... a whole



Occhialini — particle passion.

new world was revealed". After much scanning, at last they discerned an unambiguous sequence: a particle of "relatively small mass" produced a nuclear disintegration at the end of its range. This was the discovery of the  $\pi$ -meson decay.

Blackett and Powell separately won the Nobel Prize for their work on elementary particles. Both awards were made in difficult, Cold War years, and Occhialini had never made a secret of his political ideas. Pontecorvo summed it up nicely, in a famous toast: "I drink not to Beppo, but to us all: may we collaborate with him, it is a practically sure way of winning a Nobel

Prize". After a few important years in Brussels, Occhialini came back to his father's chair in Genoa and, from 1952, in Milan. Under his leadership research groups were born which largely focused on cosmic-ray studies (the 'G-stack collaboration'), but also covered the transition of elementary particle work to accelerator physics ('K-collaboration') and the beginning of space physics. In the latter field, Occhialini created in Milan a truly European school of high-energy astrophysics, out of which came two generations of active scientists.

Above all, Beppo was instrumental, together with E. Amaldi and others, in starting the European Space Research Organization, and in giving an impetus to its scientific programme from which the present-day European Space Agency still benefits. Among the first to grasp the importance of gamma-ray astronomy, he was one of the founding fathers of the COS-B project. A large fraction of its success can be traced to his scientific and human guidance.

To work with or for Beppo Occhialini has been a privilege of many scientists in diverse fields and countries the world over. For them, days of labouring under the exacting standards he set (most of all for himself) are now cherished memories. His unique style in science and culture will continue to receive the best of compliments that we can pay to it: imitation.

Giovanni Bignami

*Giovanni Bignami is in the Istituto di Fisica Cosmica, CNR, Milan, Italy.*



candidate galaxy<sup>7</sup> is NGC1275, but here the situation is complicated by its position in an X-ray-emitting cluster of galaxies, where a cooling flow of gas may be providing material for star (and perhaps cluster) formation.

The best evidence to date now comes through new Hubble Space Telescope observations<sup>1</sup> of the interacting pair of gassy galaxies NGC7252, where the good spatial resolution has shown some 40 blue objects that are far too bright to be single stars but are about the right size (10 parsec radius) to be new globular clusters. Follow-up spectroscopy<sup>2</sup> shows that at least two of these objects — and by implication the rest — are indeed young, but not extremely young, star clusters. Estimated ages are consistent with formation during the past 10<sup>9</sup> years, which is the time for which the two galaxies have been interacting. The importance of the fact that the clusters are not too young is that it indicates they are indeed gravitationally bound structures, and not just the kind of young stellar association which would break up rapidly after formation.

Because significant numbers of globular clusters can be formed in galaxy interactions, the observed density of clusters in ellipticals is no longer an argument against their formation from spirals. The predominance of ellipticals over spirals in dense clusters of galaxies now seems more natural, given the greater opportunities for interactions. The hunt will be on to identify young globulars in other systems, perhaps even in ellipticals where other signs of merger have largely dispersed.

There are alternative explanations for the large numbers of globular clusters in ellipticals<sup>8</sup> — perhaps they were just formed with more. Before rejecting such models, it must be shown that the properties of elliptical galaxies can be explained fully on the basis of galaxy interactions. But the demonstration that globular clusters can be made in interactions will greatly strengthen the case for galaxy formation and evolution by mergers. Indeed, it may mark a real shift from the general view of the isolated formation of galaxies by simple gravitational collapse. □

*M. G. Edmunds is in the Department of Physics and Astronomy, University of Wales College of Cardiff, Cardiff CF2 3YB, UK.*

## Meticulous AP-1 factors

*James Douglas Engel*

THE NF-E2 transcription factor activates transcription from a discrete set of AP-1-related sites in erythroid gene enhancers and promoters. Now it promises to do so in other cell lineages as well. As described in two papers<sup>1,2</sup> (one on page 568 of this issue<sup>2</sup>), the initial chapter in the molecular anatomy of NF-E2 concludes with the cloning of its second constituent chain, which associates with the erythroid-restricted subunit p45 (ref. 3) to constitute full NF-E2 activity. The hitherto missing subunit(s) turn out to be members of the *maf* proto-oncogene family.

To appreciate these new results, the problem needs to be seen in historical perspective. Romeo and colleagues initially described a transcription factor (NF-E2) that acts at a regulatory site required for the erythroid-specific expression of the gene encoding porphobilinogen deaminase<sup>4</sup>. For a short time, there were grumbled suspicions that this assignment might be incorrect, because the binding site for NF-E2 bears such striking similarity to the much better characterized, and far more prevalent, *cis* element recognized by the ubiquitous transcription factor AP-1. Evidence underscoring the crucial activity of this new factor emerged over the ensuing years, the sum of which strongly suggested that NF-E2 constitutes a unique activity in erythroid cells, and that it might be centrally involved in erythroid gene regulation. The molecular details of NF-E2 remained elusive, however, because the protein was expressed at much lower abundance in red blood cells than its haematopoietic predecessor, NF-E1 (now called GATA-1).

Thus a breakthrough was achieved with the heroic cloning of p45 NF-E2 last year<sup>3</sup>. But an additional wrinkle became immediately apparent: the erythroid-restricted p45 protein was incapable of binding to NF-E2 *cis* elements. Hence Andrews *et al.* speculated that a second constituent chain (p18), required for full function of NF-E2, was yet to be found (a good wager, as the conceptually translated p45 complementary DNA sequence indicated that this protein would include a leucine-zipper motif)<sup>3</sup>. Further evidence that NF-E2 is composed of two different polypeptides followed shortly thereafter<sup>5</sup>. Would the p18 partner for p45 be a Fos or Jun family member (a normal constituent of AP-1), or a completely new protein?

So a second chapter of this continuing nuclear mystery begins with the appearance of the two reports<sup>1,2</sup> identifying the second constituent chain(s) of NF-E2 as members of the *maf* proto-oncogene family. The *v-maf* oncogene was originally described as the transforming component

of the avian AS42 retrovirus, and its cellular counterparts now number five distinct, but clearly related, proteins<sup>6</sup>. The small Maf proteins (MafF, MafG and MafK) lack canonical transactivation domains, but have dimerization and DNA-binding domains, whereas large *maf* family members (*v-maf*, *mafB* (ref. 6) and *NRL* (ref. 7)) additionally encode putative transactivation domains. Thus the ubiquitous small Maf proteins have an increasingly familiar regulatory role (for example *Id/myoD* and *Max/Myc*) in development and oncogenesis as heterodimeric partners for p45, which together constitute the NF-E2 binding function in the absence of a direct contribution by Maf to NF-E2 transactivation.

The long-term implications of these findings are considerable. One intriguing possibility, specifically for those cells that express p45, can be deduced from synthesis of two earlier observations. First, although the small Maf proteins are ubiquitously expressed, there are nonetheless considerable differences in their abundance in specific cell types<sup>1,2</sup>. Second, although p45 homodimers bind poorly to NF-E2 sites<sup>2,3</sup>, small Maf homodimers acting alone suppress transcription from these sites quite effectively<sup>2</sup>. Taken together, these data suggest that an extremely sensitive and Maf-concentration-dependent switch may regulate genes through NF-E2 sites, and elicit powerful positive or negative regulatory effects by only subtle variation in the cellular concentrations of the Maf proteins<sup>2</sup> and/or p45 (for instance during globin gene 'switching'<sup>8</sup>).

These conclusions provoke other questions. Does Maf serve as an obligate repressor in most tissues at cellular NF-E2 sites (so suppressing the activity of erythroid genes in the absence of p45)? Or will discrete, tissue-restricted functional homologues of p45 shortly be identified in a number of other cell types (the retina-specific *NRL* member of the family<sup>7</sup>, for example), echoing the erythroid pattern in developmentally distinct transcriptional regulatory environments?

A further, fascinating possibility emerges from the observation that a ubiquitous intracellular factor (presumably AP-1) can also activate transcription from NF-E2 sites<sup>2</sup>. NF-E2 sequences represent a subset of the less stringent AP-1 binding-site requirement<sup>3</sup>, so one might reasonably ask whether the huge catalogue of AP-1-regulated genes is, in fact, regulated only by Jun/Fos family members. If not, might some of these presumptive AP-1 elements additionally be regulated by p45 homologue/Maf pairs

- Whitmore, B. C., Schweizer, F., Leitherer, C., Borne, K. & Robert, C. *Astr. J.* **106**, 1354–1370 (1993).
- Schweizer, F. & Seitzer, P. *Astrophys. J.* **417**, L29–L32 (1993).
- Toomre, A. in *The Evolution of Galaxies and Stellar Populations* (eds Tinsley, B. M. & Larson, R. B.) 401–416 (Yale Observatory, New Haven, 1977).
- van den Bergh, S. *Publ. Astr. Soc. Pacif.* **94**, 459–464 (1982).
- Ashman, K. M. & Zepf, S. E. *Astrophys. J.* **384**, 50–61 (1992).
- Lutz, D. *Astr. Astrophys.* **245**, 31–40 (1991).
- Richter, H. B., Crabtree, D. R., Fabian, A. C. & Lin, D. C. N. *Astr. J.* **106**, 877–885 (1993).
- West, M. J. *Mon. Not. R. astr. Soc.* **265**, 755–765 (1993).

(or Maf/Fos or Maf/Jun complexes<sup>6</sup>)? Resolution of this issue will probably lead to insights into unexpected subtleties in transcriptional regulation mediated through an extended family of AP-1 *cis* elements.

Finally, if NF-E2 p45 homologue/Maf collaborations are documented in other tissues, and these factors also act on a subset of *cis* elements regulated by AP-1, is it also possible that partial complementation of defects engendered by germline mutation of the *fos* or *jun* genes arises from outside the (anticipated) *fos* or *jun* gene families? Thus further analysis of Maf and its currently defined or prospective partners may well provide a better understanding of tissue-specific versus growth-factor-induced transcriptional re-

sponses elicited through core AP-1 *cis* elements, and of the origins of complementation of transcription-factor defects in mice with altered germ lines. □

James Douglas Engel is in the Department of Biochemistry, Molecular Biology and Cell Biology, Northwestern University, Evanston, Illinois 60208, USA.

1. Andrews, N. C. *et al.* *Proc. natn. Acad. Sci. U.S.A.* **90**, 11488–11492 (1993).
2. Igarashi, K. *et al.* *Nature* **367**, 568–572 (1994).
3. Andrews, N. C., Erdjument-Bromage, H., Davidson, M. B., Tempst, P. & Orkin, S. H. *Nature* **362**, 722–728 (1993).
4. Mignotte, V. *et al.* *Nucleic Acids Res.* **17**, 37–54 (1989).
5. Ney, P. A. *et al.* *Molec. cell. Biol.* **13**, 5604–5612 (1993).
6. Kataoka, K., Noda, M. & Nishizawa, M. *Molec. cell. Biol.* **14**, 700–712 (1994).
7. Swaroop, A. *et al.* *Proc. natn. Acad. Sci. U.S.A.* **89**, 266–270 (1992).
8. Engel, J. D. *Trends Genet.* **9**, 304–309 (1993).

self-pollen. Parallels between SI and host-pathogen interactions are not surprising<sup>3</sup>, for both involve the penetration of the 'host' by a tubular cell emanating from a spore-like structure, and in both cases the tube grows in, and sometimes between, the cell walls. Indeed it is a matter of some interest why pollen tubes fail to activate systems intended for defence against pathogens<sup>4</sup>.

First, some background. Genes linked to S-loci were initially cloned by groups led by the Nasrallahs (*Brassica*)<sup>5</sup> and Adrienne Clarke (*Nicotiana*)<sup>6</sup> in the mid-1980s. Both genes encode glycoproteins, but subsequent sequence comparisons suggested that the *Nicotiana* protein might also have RNase activity. Over the past few years Clarke's group has put forward strong circumstantial evidence implicating RNase in the inhibition of self-pollen. But difficulties in expressing heterologous S-RNases at high levels in *Nicotiana* seem to have prevented the more obvious transformation experiments.

The two sets of experiments now reported provide this elusive evidence. Lee *et al.*<sup>1</sup> have transformed self-incompatible *Petunia inflata* — a plant which is closely related to *Nicotiana*, but synthesizes massive amounts of S-RNase in the style — with an antisense S-RNase construct. Not only were the levels of the enzyme significantly reduced in these plants, but also, and more importantly, the transformants were self-compatible with pollen carrying the allele concerned. Having shown that

## SELF-POLLINATION

# Simply a social disease?

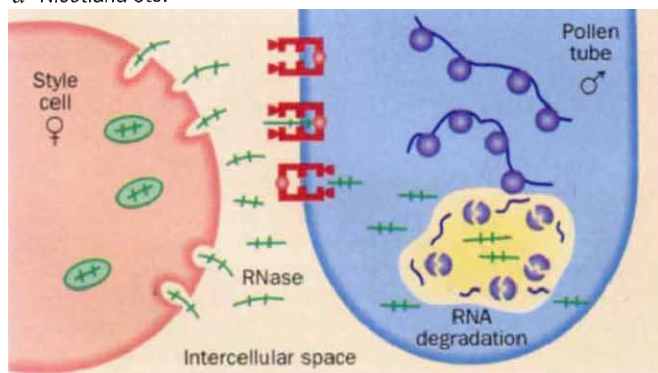
Hugh Dickinson

SELF-incompatibility (SI) systems, the physiological mechanisms that prevent self-fertilization, are helpful models for intercellular signalling in flowering plants. Our understanding of how they work takes a jump forward with the publication, on pages 560 and 563 of this issue, of reports by Lee *et al.*<sup>1</sup> and Murfett *et al.*<sup>2</sup>, who show that in one group of plants the pistil identifies and rejects self-pollen

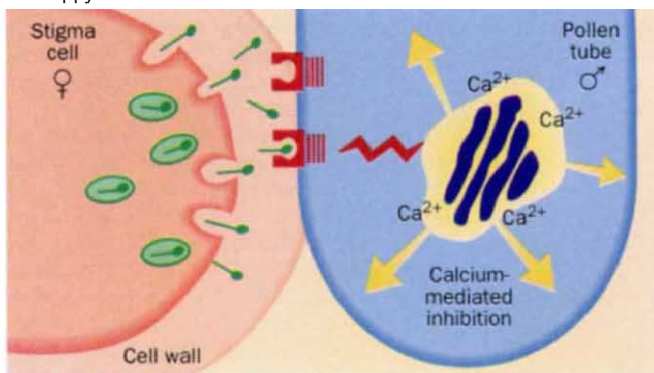
through the action of specific RNases.

Significantly, these S(incompatibility)-RNases are expressed in the transmitting tissue of the style, together with a range of enzymes which protect the vulnerable pistil from attack by pathogens. So, early in the evolution of the angiosperms, selection may have recruited a mechanism — then involved in the protection of the pistil from disease — to identify and reject

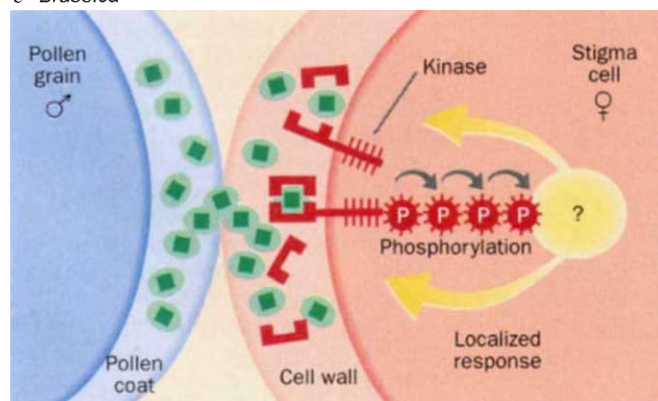
a *Nicotiana* etc.



b Poppy



c *Brassica*



The three best-characterized mechanisms of self-incompatibility in flowering plants, with the ligand depicted in green, the recognition event in red and the response in yellow. a, The RNase system found in, for example, *Nicotiana* and *Petunia*. An active RNase is secreted by stilar cells and, in self-pollinations, is recognized and internalized by receptors on the pollen-tube surface. The enzyme then presumably degrades messenger and ribosomal RNA. b, In species of poppy, a small stylar protein is presumably recognized by a receptor on the pollen-tube surface following self-pollinations. This molecule is not internalized, and it is thought that the binding of ligand and receptor generates a second message which stimulates the release of  $\text{Ca}^{2+}$  from internal stores. Increased levels of intracellular  $\text{Ca}^{2+}$  then inhibit pollen tube development. c, The situation is reversed in *Brassica* spp., where pollen molecules interact with a multicomponent female receptor system (including a kinase) on the stigma plasma membrane. Phosphorylation of intermediates then stimulates a localized response which inhibits pollen development.



the S-RNase is necessary for SI to operate, the group went on to transform SI plants (already carrying two S-alleles) with the S-RNase associated with a third, and different, S-allele. These transformants continued to reject pollen carrying the 'endogenous' S-alleles. But they also now rejected pollen carrying the heterologous allele, demonstrating that expression of the S-RNase alone is sufficient for the stigma to reject pollen of a particular S-genotype.

Similar conclusions are drawn by Murrett *et al.*<sup>2</sup>. Ingeniously overcoming the problem of low expression of heterologous S-RNase in *Nicotiana* by driving the transgene with a strong stylar promoter, they expressed an S-RNase in a self-compatible (SC) hybrid of *Nicotiana langsdorffii* and *N. alata*. The resulting transformants synthesized high levels of the heterologous S-RNase in the style, and rejected pollen from plants naturally homozygous for the S-RNase. Last year, Clarke presented data at the International Genetical Congress which, in retrospect, complement these findings. Using a *Lycopersicum* (tomato) line, which had mutated from SI to SC, her group showed that the 'S-RNase' polypeptide was present at normal levels, but was inactive enzymatically, the mutation having changed a key histidine residue at the RNase active site. These findings, which are not yet published, thus suggest that the enzymatic activity of the S-RNase is essential for the operation of SI.

Vital pieces of the jigsaw are still missing, by far the largest being the male factor. This molecule is presumably located on the pollen-tube surface, and must interact with the S-RNase so as to internalize it in self combinations. Further, we have no information on the features of the S-linked RNase itself which confer S-specificity. The strong selective advantage offered by SI (ref. 7) has resulted in the evolution of a number of different mechanisms, of which RNase is only one. The study of these other SI mechanisms has, disappointingly, failed to provide any evidence of factors common to either male or female determinants. So far, we have molecular data for three types of SI system (see figure) — first, S-RNases which are found in the Solanaceae, the Scrophulariaceae and the Rosaceae; second, a mechanism based on a smaller female protein, described to date only for a species of poppy, which has recently been shown to activate a calcium-mediated signalling system in the pollen tube<sup>8,9</sup>; and third, a very different system found in the Brassicaceae, and perhaps the Asteraceae, where the 'rejection reaction' seems to be sited in the female stigmatic cells rather than the pollen.

Study of this last system has revealed yet another link between SI and host-pathogen responses, for molecular analy-

sis of the S-locus in *Brassica oleracea* by the Nasrallahs' group has revealed the presence of sequences encoding a glycoprotein and a transmembrane serine/threonine kinase, both of which share considerable homology for any S-allele<sup>10</sup>. Both genes seem to be expressed primarily in the female. Strikingly, the S-locus receptor kinase belongs to a newly discovered family of receptor kinases, one of which acts as a disease resistance gene in tomato<sup>11</sup>.

The search for the male determinant in SI has been complicated by an elegant early theory which predicts that male and female cells synthesize similar molecules which, in incompatible combinations, dimerize to form an active agent. Thus considerable effort continues to be expended in attempts to find S-RNases and S-glycoproteins in the pollen. Tantalizing reports of such molecules continue to appear, but unequivocal evidence has not been forthcoming. However, if many or all SI mechanisms are derived from disease-resistance systems, might the elusive pollen factor be in some way related to the 'natural' ligand of these defence mechanisms? Although this may be the case for *Brassica* where a receptor kinase is involved, it is unlikely to be so in the Solanaceae, where the stylar 'defence' enzymes seem not to be associated with any signalling mechanism (although the S-RNase must have acquired one at some stage). In any event, we are currently ignorant of any ligand which binds to the 'defence' kinases.

It would be wrong to paint too depressing a picture of our knowledge of SI systems for, by plant standards, they are comparatively well understood. The genetics are clearly defined, we know a good deal about the cells involved and useful molecular data are being accumulated. The results reported in this issue now provide a mechanism by which self-pollen may be rejected. Taking a more familiar analogy, we have found the spring in the mousetrap — we need next to discover the cheese and, in the fullness of time, what makes the mouse interested in it. □

Hugh Dickinson is in the Department of Plant Sciences, University of Oxford, South Parks Road, Oxford OX1 3RB, UK.

1. Lee, H.-S., Huang, S. & Kao, T.-h. *Nature* **367**, 560–563 (1994).
2. Murrett, J., Atherton, T. L., Mou, B., Gasser, C. S. & McClure, B. A. *Nature* **367**, 563–566 (1994).
3. Hodgkin, T., Lyon, G. D. & Dickinson, H. G. *New Phytol.* **110**, 557–569 (1988).
4. Dickinson, H. G. *Nature* **364**, 573–574 (1993).
5. Nasrallah, J. B., Kao, T.-H., Goldberg, M. L. & Nasrallah, M. E. *Nature* **318**, 263–267 (1985).
6. Anderson, M. A. *et al.* *Nature* **321**, 38–44 (1986).
7. Whitehouse, H. L. K. *Ann. Bot.* **14**, 198–216 (1950).
8. Foote, H. C. C. *et al.* *Proc. natn. Acad. Sci. U. S. A.* (in the press).
9. Franklin-Tong, V. E., Ride, J. P., Read, N. D., Trewavas, A. J. & Franklin, F. C. H. *Plant J.* **4**, 163–177 (1993).
10. Dzelkals, V. A., Nasrallah, J. B. & Nasrallah, M. E. *Dev. Biol.* **153**, 70–82 (1992).
11. Martin, G. B. *et al.* *Science* **262**, 1432–1436 (1993).

## Elastic gas traps

THE standard gas cylinder, says Daedalus, is an absurdity — 50 kg of steel wrapped around a mere 10 kg of gas. He recalls that active charcoal and zeolite 'molecular sieves' can take up almost their own weight of gas in their molecule-sized pores. Sadly, only strong heat can get it out again. He is now inventing a novel elastic microporous solid that is more like a molecular sponge.

He points out that some polymers go opaque under stress (indented-letter labelling tape works this way). Tension opens innumerable microvoids in their structure, which scatter the light. Relax the stress and the voids close up; the polymer clears again. DREADCO's chemists are now taking this effect to extremes. Their goal, DREADCO's 'Gasponge', will expand under tension into an open-cell microfoam with pores well sized to accept gas molecules.

A lot of force will be needed to dilate Gasponge. Every gram will develop hundreds of square metres of internal surface, giving the polymer an energy density approaching that of TNT. It will relax that energy by absorbing gas avidly into its internal pores, releasing the energy of absorption as latent heat. Once loaded, a Gasponge sachet will hold its charge of gas indefinitely. To get the gas out again, simply squeeze it in a vice, when the voids will close up and expel the trapped gas. Heavy pressure will be needed, and the sample will cool strongly by the uptake of latent heat.

Gasponge will hold and dispense most gases, from the hazardous chlorine and ammonia to safer ones such as carbon dioxide and butane. The one gas which presents problems is, sadly, oxygen. Like oxygen-loaded charcoal, an oxygen-loaded hydrocarbon polymer would be dangerously inflammable and possibly highly explosive. For welders, glassblowers, anaesthetists, divers and other oxygen users, Daedalus hopes to devise a special non-inflammable Gasponge based on silicone or fluorocarbon rubber.

Many other uses should open up for Gasponge. By squeezing and relaxing a sample, its latent heat flow could be exploited in a simple but powerful refrigerator. As a gas trap, it could remove exhaust fumes and foul air from enclosed spaces. A small compressible Gasponge could make a neat aerosol propellant. And an oxygenated Gasponge with a sharp melting point should expand forcefully if heated to that point. This novel smokeless non-burning explosive would be ideal for mining and tunnelling. The oxygen it released would even help the ventilation. David Jones

# Purification of nanotubes

**SIR** — Carbon nanotubes can now be produced in large quantities, either as multi-shell<sup>1,2</sup> or as single-layer<sup>3,4</sup> tubes. But together with the growth of the tubular material, other carbonaceous materials are always formed and their presence has hampered the accurate characterization of the bulk properties of nanotubes. We report here a simple method to purify nanotubes which should be useful for further studies.

In the case of the arc-discharge process yielding multi-shelled nanotubes<sup>1,2</sup>, nanoparticles are the main by-product and compose about one-third of the sample in the best preparations. To purify the nanotubes and eliminate the nanoparticles, we first tried various standard techniques such as filtration, chromatography and centrifugation of sonicated solutions of the raw material. Although these techniques will separate to some extent large clumps of material from individual nanotubes and nanoparticles, they did not separate the nanotubes from the nanoparticles.

Recently, we and others reported that nanotubes could be opened and thinned through oxidation<sup>5,6</sup>. The significant difference in the oxidation reaction rates of the carbon nanotube caps as compared to the cylindrical surfaces results in the production of open nanotubes during oxidation. The enhanced reactivity of the caps

results either from the strong local curvature and imperfect geometry and/or from the presence of five-membered carbon rings as discussed previously<sup>6</sup>. The structure (number of pentagons in a closed shell) and geometry of nanoparticles very much resembles that of the tube caps, although they are comparatively larger<sup>7</sup>. So nanoparticles should be similarly consumed by oxidation although at a slower rate than the highly curved (strained) nanotube tips.

Once the nanotube caps are destroyed, the tubes essentially consist of a perfect hexagonal carbon network and the reaction proceeds slowly. But in the case of nanoparticles all the concentric layers are similar in structure and etching of the layers should occur at a relatively constant rate.

The above suggests that at some point during oxidation, all the caps and nanoparticles should be destroyed, leaving only open nanometre-sized cylinders of carbon in the sample. We observe just this if the oxidation is allowed to proceed for long durations. The figure shows the dramatic improvement in the nanotube-to-nanoparticle ratio after this treatment. Note that to remove all the nanoparticles from the samples, one has to oxidize more than 99% of the material. When only about 95% of the sample is oxidized, just 10–20% of the remaining

material comprises pure nanotubes, but this still represents a substantial reduction in the proportion of nanoparticles in the total sample. In samples oxidized to a degree of less than 85%, no change in the ratio of nanoparticles is apparent relative to the original sample.

In the final purified sample, all the tubes are open and the partly removed outer layers contain reactive edges<sup>5</sup>. The average aspect ratios have decreased but not changed significantly (most are >100 but there are a few open cylinders with aspect ratios >20) as the tubes were originally much longer than the particles. Why more than 99% of the weight has to be burnt off to produce pure nanotubes is not clear. One possible explanation is that the oxidative species are not able to reach evenly all the parts of the sample because the ground-up nanotubes and nanoparticles are intertwined. We have tested this possibility by better dispersal of the samples before oxidation and by using flowing oxygen during the reaction. But so far we have not noticed any significant change. A more likely reason why the sample must be consumed significantly for purification is that the difference in reactivity between the nanotubes and the nanoparticles is small.

Whether this technique can be used to purify single-shell nanotubes, where soot and catalysts are major by-products<sup>3,4</sup>, will depend very much on their relative reactivity. The single-shell nanotubes, having diameters approaching those of fullerenes, might be much more reactive, as is the case with C<sub>60</sub> (refs 6,8).

**T. W. Ebbesen**

**P. M. Ajayan†**

**H. Hiura**

**K. Tanigaki**

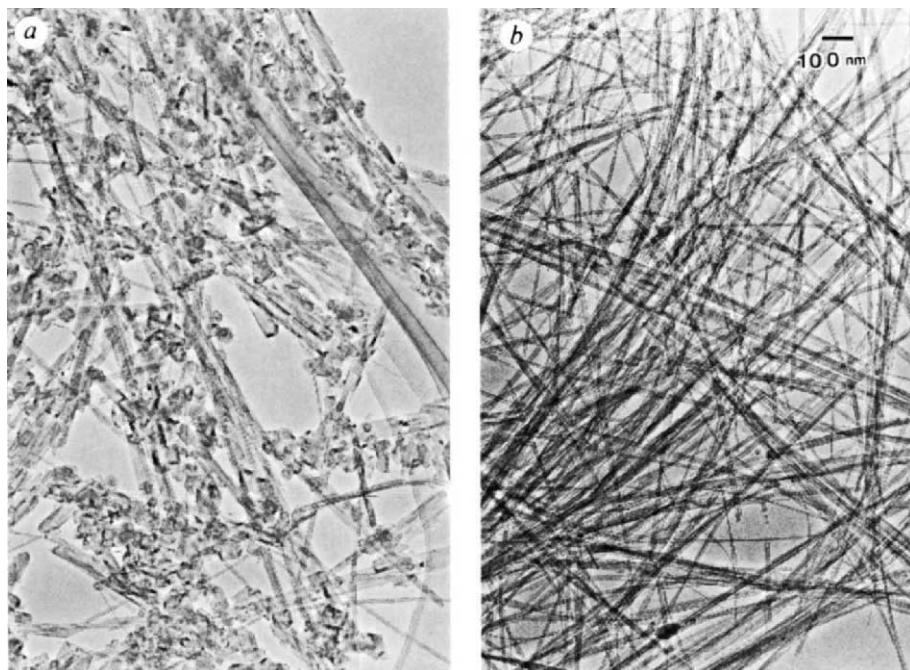
*Fundamental Research Laboratory,  
NEC Corporation, 34 Miyukigaoka,  
Tsukuba 305, Japan*

†Present address: Laboratoire de Physique des Solides, University of Paris-Sud, Bat. 510, Orsay 91405, France.

1. Ebbesen, T. W. & Ajayan, P. M. *Nature* **358**, 220–222 (1992).
2. Ebbesen, T. W. *et al. Chem. phys. Lett.* **209**, 83–90 (1993).
3. Iijima, S. & Ichihashi, T. *Nature* **363**, 603–605 (1993).
4. Bethune, D. S. *et al. Nature* **363**, 605–607 (1993).
5. Tsang, S. C., Harris, P. F. J. & Green, M. L. H. *Nature* **362**, 520–522 (1993).
6. Ajayan, P. M. *et al. Nature* **362**, 522–525 (1993).
7. Ajayan, P. M., Ichihashi, T. & Iijima, S. *Chem. phys. Lett.* **202**, 384–388 (1993).
8. Pang, L. S. K., Saxby, J. D. & Chatfield, S. P. *J. phys. Chem.* **97**, 6941–42 (1993).

## Scientific Correspondence

Scientific Correspondence is intended to provide a forum in which readers may raise points of a scientific character. They need not arise out of anything published in *Nature*. In any case, priority will be given to letters of fewer than 500 words and five references.



Low-magnification transmission electron microscope images of *a*, a typical nanotube sample produced by the carbon-arc method; and *b*, a purified open carbon nanotube sample. The purified sample was prepared as follows. Ground raw nanotube sample taken from the inner core of the carbon arc deposit<sup>1,2</sup> was placed in an oven and then the temperature was raised to 750 °C (in air or oxygen). After about 30 min, about 1% remained with the quality shown in *b*. In most tubes, the length/diameter ratios exceed 100, although there are some open cylinders with the ratio as small as 20.

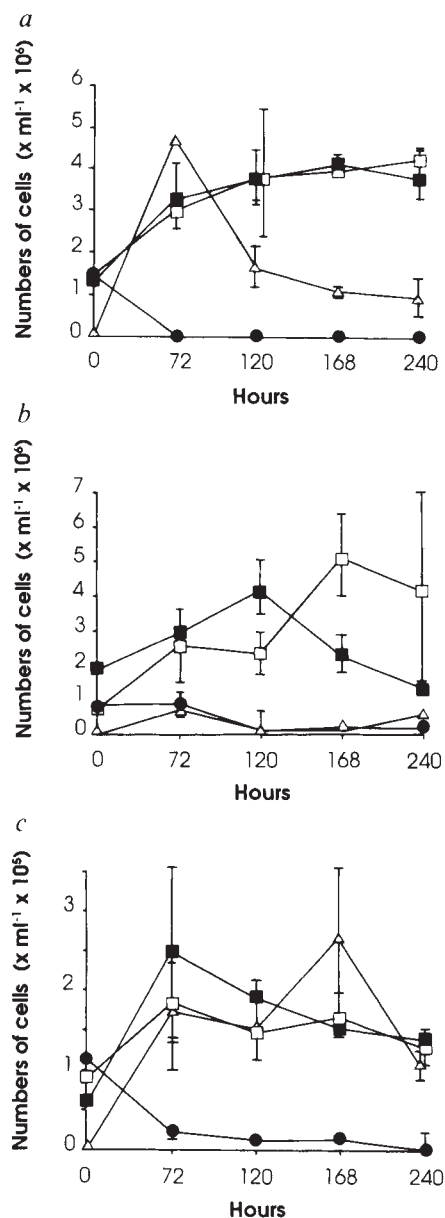


# Diatom kills by flagellates

**SIR** — The idea of 'distinct trophic levels' (see ref. 1), with producers, consumers and predictable food chains, has been challenged<sup>2,3</sup>. Here we show that the common bacterivore flagellate *Cafeteria roenbergensis* caused mass mortality in cultures of various marine diatoms, thereby supporting increased growth of bacteria feeding on the decaying algae. When inoculated with diatoms, algae flocculated into larger aggregates crowded with flagellates within a few days, causing the sinking and senescence of the cultures. This 'gardening' mechanism may be important in promoting sudden collapses of diatom blooms, formation of marine 'snow' and sequestration of carbon from a grazer food chain to a detritus food chain.

We investigated this phenomenon after repeatedly observing flocculation and subsequent dieback of *Skeletonema costatum* in cultures contaminated with the flagellate *C. roenbergensis*. We added *C. roenbergensis* to the common marine diatoms *S. costatum*, *Chaetoceros socialis*, *Thalassiosira pseudonana* and the cryptomonad *Rhodomonas baltica*. At 24-hour intervals for 7 days, we counted numbers of bacteria, *C. roenbergensis* (identified by N. Vørs) and algae by using both Utermöhl counting chambers and a combined DAPI and Primulin staining method<sup>4</sup>. We observed that *C. roenbergensis* attached to the surface of *S. costatum* after 24 hours and that after 48 hours, aggregates apparently generated by the flagellates were beginning to decay. When the aggregates began to decay, they were heavily colonized by bacteria which, again, were grazed by *C. roenbergensis*. Complete algal dieback was observed during the following 5 days. The same pattern was observed for the other species of diatoms investigated, but no effect was observed for the cryptomonad.

To test whether extra cellular cellulytic enzymes were involved in the formation of aggregates, we inoculated algae with filtrate from cultures with decaying algae and a dense population of the flagellate (collapsed culture). Fresh 10-ml batch cultures of diatoms were added to 1 litre of 0.2 µm filtrate (removing flagellates) from the collapsed cultures and followed for 10 days. This procedure also served as a test for pathogenic viruses which could cause corresponding collapses in algal populations<sup>5</sup>, as viruses would have been present in the filtrate. In parallel, we inoculated fresh cultures of algae with *C. roenbergensis*. Fresh cultures of algae without added filtrate or flagellates served as controls. Each culture was counted at 24-hour intervals, and we observed the



same effect. Aggregates of sinking algae were formed soon after addition of *C. roenbergensis*, while there were no such effects in the algae exposed to filtrate where *C. roenbergensis* had previously been present.

Diatoms are dominant among marine phytoplankton in most northern waters, forming frequent spring blooms which can suddenly collapse. Such blooms are recognized sources of marine 'snow'. Because grazing losses are low, significant amounts of carbon from sinking blooms are sequestered to the sediments<sup>6</sup>. Protists, mainly small heterotrophic flagellates, have been

described as the pioneer organisms that colonize marine snow<sup>7</sup>, and *C. roenbergensis* is a common heterotroph nanoflagellate reported to be a common marine snow protist throughout the world<sup>8</sup>.

We suggest that marine protists not only colonize marine snow but play an active part in the formation of such aggregates. This may be an important mechanism promoting termination of diatom blooms. The phenomenon of flagellates controlling organic aggregates has been observed in sewage plants, which rely on microbial flocculation involving protists. The same phenomenon may be at work here, suggesting that flagellates have a significant impact on primary production, energy budgets and carbon flux by 'triggering' a sink from energy generated by photosynthesis in the euphotic zone.

**Kari Nygaard**

**Dag O. Hessen**

NIVA, PO Box 173 Kjelsås,  
N-0411 Oslo, Norway

1. Lindeman, R. I. *Ecology* **23**, 399–418 (1942).
2. Bird, D. F. & Kaliff, J. V. *Science* **231**, 493–494 (1986).
3. Polis, G. *Am. Nat.* **138**, 123–155 (1992).
4. Nygaard, K. N., Børsheim, K. Y. & Thingstad, T. F. *Mar. Ecol. Prog. Ser.* **44**, 159–165 (1988).
5. Suttle, C. A. *et al.* *Nature* **347**, 467–469 (1990).
6. Wassmann, P. *Mar. Poll. Bull.* **21**, 183–187 (1990).
7. Caron, D. A. in *The Biology of Free-living Flagellates* (Eds Patterson, D. J. & Larsen, J.) 77–92 (Oxford University Press, Oxford, 1991).
8. Patterson, D. J. *et al.* *J. Mar. Biol. Ass.* **73**, 67–95 (1993).

## Sex and longevity

**SIR** — Nieschlag *et al.*<sup>1</sup> say that "...the difference in life expectancy between men and women may be related to another factor, probably Y-chromosome-mediated, rather than to testosterone and the testes." We agree that testosterone and the testes are not related to sex differences in life expectancy simply because boys die more often at any age<sup>2</sup>. We also agree that it is reasonable to assume that gender gap in human longevity is "probably Y-chromosome mediated" simply because male sex itself is linked to the Y-chromosome in humans.

There is, however, one important observation which should be taken into account: in other mammalian species (such as rats, mice and hamsters) where male sex is also linked to Y-chromosome, males generally do not live shorter than females<sup>2</sup>. Thus, Y-chromosome *per se* cannot be a general explanation for sex differences in life span.

**Leonid A. Gavrilov**

**Natalia S. Gavrilova**

A. N. Belozersky Institute,  
Moscow State University,  
Moscow 119899, Russia

1. Nieschlag, E., Nieschlag, S. & Behre, H. M. *Nature* **366**, 215 (1993).
2. Gavrilov, L. A. & Gavrilova, N. S. *The Biology of Life Span: A Quantitative Approach* (Harwood Academic, Chur, London, 1991).

# Chomsky's revolution

Neil Smith

**The Linguistics Wars.** By Randy Allen Harris. Oxford University Press: 1993. Pp. 356. £22, \$30.

NOAM Chomsky's position in the history of ideas is comparable to that of Darwin or Descartes. In this century his peers in influence are the unlikely trio of Einstein, Picasso and Freud, with each of whom he has something in common. Like Darwin and Descartes, Chomsky has redefined our understanding of ourselves as humans; like Freud — but with added intellectual rigour — he has revolutionized our view of the mind; like Einstein, he blends intense scientific creativity with radical political activism; like Picasso, he has overturned and replaced his own established systems with startling frequency. That a book on the history of linguistics should be reviewed in *Nature* is ultimately due to the fact that Chomsky's work has brought the study of language from the impressionism of the humanities into the scientific fold.

Apart from its direct effect on linguistics, Chomsky's work has had a major influence on philosophy and psychology, and a minor but not insignificant effect on a range of disciplines from anthropology to mathematics, from education to literary criticism. To understand this pervasive influence presupposes a grasp of the defining characteristics of Chomsky's programme, which combines mentalism, rationalism, genetic determination and the psychological reality of the constructs he postulates. Although behaviourists eschewed any appeal to the mental as being irredeemably unscientific, contemporary psychology — of which linguistics forms a part — is based firmly on the causal efficacy of beliefs and desires. Moreover, what underpins this mentalism is a version of Cartesian rationalism that ascribes massive innate cognitive structure to the neonate. Appealing to arguments both from the poverty of the stimulus and universal properties of natural (human) languages, Chomsky has reinstated an epistemology that had seemed extinct, and in doing so he has produced explanatory theories that are making possible an understanding of the nature of language and, most importantly, how language is acquired.

One core idea of this current work, known as 'principles and parameters theory', is that humans are innately endowed with a set of universal principles

that constrain the development of language. That is, the putative hypothesis space of the infant language-learner includes so few possibilities that the task of language acquisition is dramatically simplified. For instance, the principle of 'structure dependence' ensures that no child will ever entertain the hypothesis that one way of relating sentences is to reverse their word order, so that the question (or negative, or future tense) of

IMAGE  
UNAVAILABLE  
FOR COPYRIGHT  
REASONS

**"Chomsky's work has brought the study of language from the impressionism of the humanities into the scientific fold."**

'Chomsky demolished behaviourist psychology' might be 'psychology behaviourist demolished Chomsky'. Some of these universal principles are 'parameterized', allowing limited and antecedently specified variation from language to language, reducing the child's task in learning the possible structure of its first language to that of selecting the values exhibited by the ambient language. As in the case of immunology, linguistics has graduated in a generation from an instructive to a selective mode.

The value of Harris's outstanding, if overly ornate, book is that it shows how we have come to where we are. He documents in meticulous detail, with great sensitivity and unswerving impartiality, how Chomsky's early theories captured the imagination of the new generation of

cognitive scientists and resulted in the overthrow of Bloomfieldian linguistics and behaviourist psychology more generally. He explains the elegance of 'deep structure' and the power of Chomsky's conception of language as expounded in the 'standard theory', and then shows how a disparate group of young scholars, the generative semanticists, effectively hijacked the fledgling theory and developed it in ways so radical that Chomsky soon came to be seen as a reactionary fighting a rearguard action against the forces of progress.

The sequel to this apparent decline was remarkable. After a decade of academic savagery in which the discipline was severely factionalized, it was Chomsky rather than the young Turks who emerged victorious. His success was due in part to the

Steve Pyke/Network

awesome power of his rhetoric, but much more to the equally awesome power of his positive new ideas. While his rivals appeared to be floundering in a morass of new and unconstrained rule types, Chomsky developed a set of conditions on rules of grammar, which eliminated the perceived excesses of the theory, and which have culminated in the 'minimalist programme'. The success of the principles and parameters framework, allowing at last for a possible explanation of the mystery of language acquisition, has excited a new generation of cognitivists just as his work of the 1950s and 1960s had excited the old.

Harris has captured the flavour and the fervour of the debates to perfection. His account of these battles is of interest because it sheds light on the emergence and development of ideas now seen to be seminal. Just as details of the manoeuvres in the battle of Borodino are important for what they reveal about Napoleon, the individual skirmishes reported here are valuable for what they reveal about Chomsky and for casting light on a crucial episode in the history of ideas. Much recent historiography of linguistics is blighted by misrepresentation or bigotry, but as an eyewitness of the events depicted, I can vouch for the accuracy and fairness of Harris's dissection. He reports that both Chomsky and George Lakoff (the most visible and voluble of the generative semanticists) are, despite all his endeavours, in 'violent disagreement' with the substance of the book. They should rather be content. Harris has achieved the near impossible: being fair to both sides in a civil war. □

Neil Smith is in the Department of Phonetics and Linguistics, University College London, Gower Street, London WC1E 6BT, UK.



## The future of science

John Ziman

**Science and Anti-Science.** By Gerald Holton. *Harvard University Press*: 1993. Pp. 203. £19.95, \$29.95.

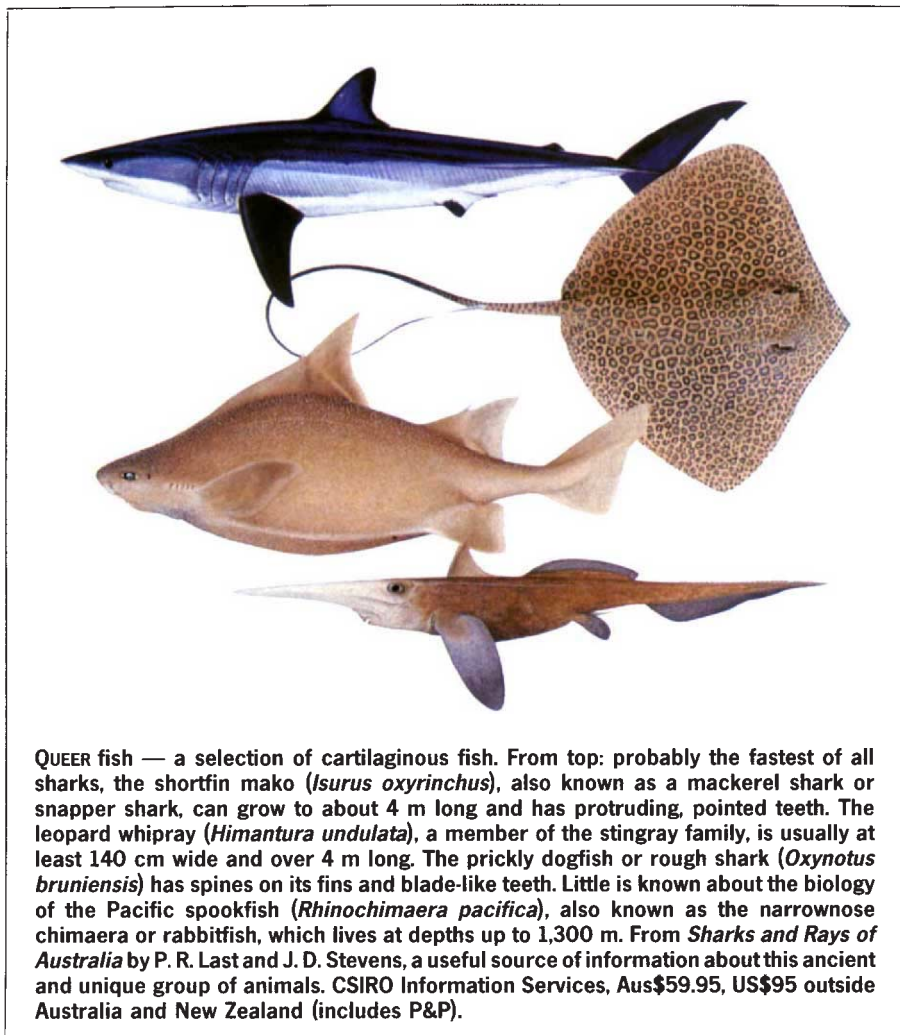
FOR the first four chapters, this seems a gentle, rather academic book, tracing subtle cultural currents in the history of modern science. That most philosophical of physicists, Ernst Mach, is identified as a source of inspiration for more general ('logical') positivism and also, unexpectedly, as an opponent of apparently positivistic relativity theory. The seminal papers on special relativity and quantum theory are unmasked as rhetorical dramas of supposed supporters and imagined opponents. Thomas Jefferson's enthusiasm for geographical exploration is interpreted as 'strategic research', seeking basic knowledge primarily for potential practical benefit.

Chapter 5, however, alerts the reader to a darkening cloud. Oswald Spengler was not the first European writer to portray the thought style of science as an internal cancer of our culture, but in 1918 his visionary bestseller, *The Decline of the West*, was received seriously by many intellectuals. Spengler's gradiose cyclical theory of history was not new, and now seems totally implausible. Nevertheless, his trumpet blast against science and technology still echoes on, and has been amplified and re-broadcast by many subsequent prophets of decadence and doom.

Holton typecasts Spengler as the progenitor or harbinger of the "anti-science phenomenon", which he dissects energetically and passionately in his final chapter. For most readers of *Nature*, the interest of the whole book is concentrated in these last 40 pages, which can be read separately from the rest.

In Holton's view, the most malignant of the many manifestations of anti-science is "the type of pseudo-scientific nonsense that manages to pass itself off as an 'alternative science' and does so in the service of political ambition". On the one hand, the high level of general scientific illiteracy in the United States and other advanced countries is a potential source of erroneous policy and eventual social instability. On the other, there are reassuring factors, such as continued popular enchantment with high technology, and the absence of generalized conflict between organized science and conventional sectarian religion. General recognition of these factors is the basis for numerous campaigns to improve 'public understanding of science'.

Holton's analysis goes deeper. He sees



**QUEER fish** — a selection of cartilaginous fish. From top: probably the fastest of all sharks, the shortfin mako (*Isurus oxyrinchus*), also known as a mackerel shark or snapper shark, can grow to about 4 m long and has protruding, pointed teeth. The leopard whipray (*Himantura undulata*), a member of the stingray family, is usually at least 140 cm wide and over 4 m long. The prickly dogfish or rough shark (*Oxynotus bruniensis*) has spines on its fins and blade-like teeth. Little is known about the biology of the Pacific spookfish (*Rhinochimaera pacifica*), also known as the narrownose chimaera or rabbitfish, which lives at depths up to 1,300 m. From *Sharks and Rays of Australia* by P. R. Last and J. D. Stevens, a useful source of information about this ancient and unique group of animals. CSIRO Information Services, Aus\$59.95, US\$95 outside Australia and New Zealand (includes P&P).

anti-science as a movement that would delegitimize science as the progressive core of modern culture. A number of different enemies are hammering at its gates. Certain philosophers seem intent on undermining the special status of scientific knowledge: some cultural critics follow the example of Spengler, and more recently of Arthur Koestler, in attacking science on a broader front: most counter-cultural, new age and Eastern mystical devotees deplore the concept of objectively reachable data: many radical feminists object to the 'androcentrism' of established science.

The particular areas of criticism — inhumane technology, environmental devastation, technocratic authoritarianism — are deplorable, but separately corrigible. Taken together, however, these attacks threaten to topple the 'world picture' that has developed around science over three centuries. Holton's succinct explication of this complex concept is too rich with compelling insights to be summed up in a few lines, but his final conclusion is simple: science itself is not obviously exempt from 'the potential for sudden and catastrophic shifts of the total world view'.

The disastrous consequence would be the supersession of 'modernism' as the driving force of our civilization. This again is a very complex concept, about which philosophers, anthropologists, sociologists and historians debate inconclusively. Is it not just another term, for example, for the common social and behavioural features of a small number of economically advanced political democracies? Holton notes this debate, but extricates himself from it by concisely characterizing the 'modern' personality in terms of a page-long list of cognitive attitudes, such as "High place for 'objectivity'", "Skepticism with respect to authority" and so on. He is thus in a position to define anti-science through a corresponding list of contrary attitudes: "Subjective, not objective", "Faith-based" and so on.

Up to this point, with due regard by both of us to numerous qualifications, my old friend Gerald Holton and I would travel together, and fight side by side for reason and justice. But I am not carried further into the gloom by the conviction that 'anti-science' constitutes a coherent world view. The "Creation Science" that causes him so much concern does not seem an intellectual partner of, say, tradi-

tional Hindu or Islamic fundamentalism, let alone the neo-Stalinism and other Fascistic movements likely to erupt in the former Soviet sphere.

In effect, Holton unconsciously adopts a Spenglerian thema of cyclical movement between two cultural poles. A more evolutionary analysis would open our imagination to the likely emergence of quite unprecedented social and psychological formations, as novel in their own diverse ways as science itself — and perhaps as beneficial. Such a 'post-modern' conclusion need not breed complacency. On the contrary, it heartens us to press on along new paths without anguish over our inability merely to stand firm or turn back. But this is only one opinion stimulated by this extraordinarily thoughtful, penetrating and wise tract for everyone concerned about the future of science. □

John Ziman is Emeritus Professor of Physics of the University of Bristol. He is at 27 Little London Green, Oakley, Aylesbury, Bucks HP18 9QL, UK.

## Licence to kill

Peter Singer

**Monkey Business: The Disturbing Case that Launched the Animal Rights Movement.** By Kathy Snow Guillermo. *National Press Books*: 1993. Pp. 254. \$23.95; UK distributor: Gazelle, £21.50.

**In the Name of Science: Issues in Responsible Animal Experimentation.** By F. Barbara Orlans. *Oxford University Press*: 1993. Pp. 297. \$39.95, £27.50.

FOR anyone wanting to know why the issue of animal experimentation has become so polarized in the United States, the simultaneous appearance of *Monkey Business* and *In the Name of Science* is an opportunity for enlightenment. Both books are written by North Americans who think that there is something wrong with the way animals are treated in research, and both describe developments in this area that have taken place within the past 10 or 15 years. There the similarity ends. Kathy Snow Guillermo and F. Barbara Orlans discuss different events and represent very different approaches to the issue.

*Monkey Business* traces the history of one celebrated case: the prosecution in 1981 of Edward Taub for cruelty to 16 crab-eating macaques and one rhesus monkey at the Institute for Behavioral Research in Silver Spring, Maryland. For the movement in the United States against animal experimentation, the case set many historical firsts: the first police raid on the laboratory of a scientist suspected of breaching laws relating to cruelty to animals; the first criminal prosecution and

conviction of a researcher for cruelty to animals (the conviction was later set aside, on the scarcely reassuring ground that Maryland animal protection laws did not apply to experimenters receiving federal grants); and the first instance in which the National Institutes of Health (NIH) withdrew financial support because of non-compliance with standards of treatment relating to animal welfare. The book also covers the long legal and political battle over the fate of the surviving monkeys, a battle which is still not over, although most of the monkeys in dispute have now died.

To suggest, as the subtitle of *Monkey Business* does, that the case of the 'Silver Spring monkeys' launched the animal rights movement, is to take a North American, perhaps even a Washingtonian, perspective. By 1981 there were already growing organizations in Britain and Australia, for example, working quite effectively for animal rights, or animal liberation, and the case had little impact in those countries. Even in the United States, the first animal rights campaign to succeed in stopping a series of experiments was directed against the American Museum of Natural History in New York in 1976–77. Nevertheless, there is no doubt that the Taub case propelled the issue to new heights of publicity, and transformed People for the Ethical Treatment of Animals — the organization behind the prosecution — from a two-person operation to the largest animal rights group in the world, with 400,000 members, 70 employees and a budget of \$10 million.

Perhaps the most disturbing aspect of the case is the way in which the US scientific community rallied to support Taub. On the evidence produced in court, there can be little doubt that, legal technicalities aside, and whatever one thinks of the nature and value of the research itself, Taub failed to provide basic veterinary care to the monkeys in his laboratory. Referring to the kind of surgery performed by Taub, and the observed consequences, officials at NIH were later to admit that "... fractures, dislocations, lacerations, punctures, contusions, and abrasions with accompanying infection, acute and chronic inflammation, and necrosis are not the inevitable consequences of deafferentation". The NIH also found that the laboratory's animal care and use committee "appeared unfamiliar" with

the substance and purpose of the NIH guidelines, compliance with which it was supposed to ensure. Yet the American Psychological Association gave Taub \$10,000 towards his legal fees and, together with several other scientific bodies, sent witnesses to testify on Taub's behalf at his appeal. At Taub's trial, veterinarians gave evidence in his favour, basing their professional diagnosis on photographs — because they had never visited the laboratory while the monkeys were there — and refusing even to condemn the fact that, despite the fractures, dislocations, infections and so on, the laboratory's only consulting veterinarian had never been called upon to examine or



The legal and political battle over the fate of the surviving monkeys is still not over.

treat any of the surgically crippled monkeys. A decade later, one of these veterinarians was appointed director of the Office of Animal Research Issues at the National Institute of Mental Health. In 1986, Taub was elected a fellow of the American Association for the Advancement of Science. With such attitudes among US scientists, it is not surprising that the debate about animal experimentation in the United States is more polarized than in Europe or Australia.

Barbara Orlans shows a better way forward. A physiologist and former animal experimenter herself, Orlans is now attached to the Kennedy Institute of Ethics at Georgetown University in Washington DC, where she studies attempts to regulate the use of animals in research. *In the Name of Science* describes the development of the regulatory approach, again with an American focus,



but at least with an awareness that US researchers can learn from what has been done in other countries. The book is full of helpful, detailed information about animal abuse in laboratories, and systems of review and regulation designed to avoid such abuse. But Orlans also gives depressing evidence of the polarization of the US debate, with lobby groups funded by the animal research industry opposing every attempt by the animal rights movement to achieve the most moderate gains in animal welfare. These lobby groups have even opposed attempts to end the utterly indefensible exclusion of rats, mice and birds from the federal regulations governing animal experimentation in the United States.

One can only wish that every US scientist involved in research on animals would read this book and reflect on its central message. Although it is true that the ethical principle of equal consideration of

interests, on which the animal liberation movement is based, is fundamentally incompatible with the use of sentient creatures as tools for laboratory research, there is still some common ground between realistic members of the animal rights movement, who want to reduce animal suffering sooner rather than later, and those scientists who can see that a failure to give serious consideration to the interests of animals is both unethical and harmful to the reputation of science itself. If there were more scientists who, like Orlans, were prepared frankly to acknowledge abuses and to work with the animal rights movement to try to eliminate them, a constructive dialogue of the kind that has taken place in other countries might also be possible in the United States. □

*Peter Singer is in the Centre for Human Bioethics, Monash University, Melbourne, Victoria 3168, Australia.*

## Ancient Eastern remedies

*Leslie Crombie*

**Traditional Medicine.** Edited by Biswapati Mukherjee. *International Science Publisher: 1993. Pp. 403. \$45.*

SOME scientific books are elegantly constructed around a central theme so beautifully expounded that a reader is made to feel that there is little else to be added. Others are something of a hotchpotch, having the compensating feature that there are things undone, unexplained or only partially finished on which it would be very interesting to do more work. This book, the record of an international seminar held in Calcutta in November 1992, comes into this second class because it covers many diverse areas.

The organizers and the authors have taken a catholic view of what is implied by traditional medicine, so there is a good account of rather nontraditional immunosuppressants of microbiological origin such as the cyclosporins, FK506 and rapamycin. There is also an interesting account of steviol glycosides, from the plant *Stevia rebaudiana*, as sweetening agents and their modification by enzymic transglycosylation. With the withdrawal of saccharin and cyclamate, the Japanese food industries have had to import some 1,500 tons of *Stevia* leaves each year.

Natural aroma chemicals as well as marine natural products, antidiabetics and prostaglandin inhibitors are discussed by other authors.

But the main thrust of the book lies in traditional medicine, especially as practised in north India and Bangladesh, where two such systems run side by side.

IMAGE  
UNAVAILABLE  
FOR COPYRIGHT  
REASONS

*A street vendor selling traditional medicines, Arambul, India.*

One is the Ayurvedic system, based on the Vedas, the oldest scripture of the Aryan age which dates from around the second millennium BC; the other is the ancient but later Unāni system of Greek and Arab medicine as developed by Arab and Mus-

lim scholars. A Westerner, accustomed to much recirculated material, will appreciate the authenticity of the contributions in this area. Sri Lankan practice is also described, along with unusual accounts of Tibetan and Turkish traditional medicine. The international dimension of the book, however, is not strong. There is very little on the important subject of Chinese traditional medicine, almost nothing on African practices, and too little on South American remedies.

The topic of traditional medicine tends to excite rivalry between advocates of traditional and Western systems. Such partisanship is largely prejudice: both use chemical compounds to effect treatment. In former times, or in undeveloped areas of the world, plants had to be the major source of such chemicals. The developed world has at its command an enormous range of synthesized chemical compounds, and can modulate these structures to achieve optimum medicinal effect. Both plants and chemical synthesis can afford poisonous as well as benign compounds: synergistic effects can sometimes be produced by mixtures in both cases. The big differences are in cost, availability and effectiveness. Poor communities cannot afford Western drugs, and bad communications and primitive conditions can hinder availability even when they can be paid for. At the same time, the local doctor or prescriber who understands and speaks the language of the patient will often be trained only in

The Hutchinson Library

indigenous medicine. Thus 75 per cent of the rural population of Bangladesh go to village doctors who use Unāni or Ayurvedic medicine. There are of course many deficiencies in traditional medicine, ranging from difficulties in the supply of plants that do not grow locally to the standardization of doses. Some plants are probably useless for the condition for which they are prescribed, but traditional medicine nonetheless helps and comforts those who would otherwise have nothing.

This book deserves a place in phytochemical and ecological libraries and in those of departments of pharmacy and medicine. There is much of general interest, and the heterogeneous nature of the articles, with their mixed interests and disciplines, should stimulate new research. □

*Leslie Crombie is in the Department of Chemistry, University of Nottingham, Nottingham NG7 2RD, UK.*

# ATP-dependent nucleosome disruption at a heat-shock promoter mediated by binding of GAGA transcription factor

Toshio Tsukiyama, Peter B. Becker\* & Carl Wu†

Laboratory of Biochemistry, National Cancer Institute, National Institutes of Health, Building 37, Room 4C-09, Bethesda, Maryland 20892, USA

**Genetic control elements are usually situated in local regions of chromatin that are hypersensitive to structural probes such as DNase I. We have reconstructed the chromatin structure of the *hsp70* promoter using an *in vitro* nucleosome assembly system. Binding of the GAGA transcription factor on existing nucleosomes leads to nucleosome disruption, DNase I hypersensitivity at the TATA box and heat-shock elements, and rearrangement of adjacent nucleosomes. ATP hydrolysis facilitates this process, suggesting that an energy-dependent pathway is involved in chromatin remodelling.**

THE role of nucleosomes as general repressors of transcriptional initiation in eukaryotic cells is well established<sup>1-3</sup>, but the mechanisms by which transcription factors, enhancer proteins and RNA polymerases gain access to target sequences in chromatin remain poorly understood. *In vivo*, genetic control elements are usually situated in accessible, nuclease-hypersensitive sites that punctuate the orderly array of nucleosomes on the chromatin fibre<sup>4,5</sup>. Generation of these accessible regions in chromatin seems to be a prerequisite for the formation of an active transcription complex, and may involve a class of transcription factors whose binding alters the stability of underlying or adjacent nucleosomes<sup>6,7</sup>.

We have developed an *in vitro* assay for the establishment of a nucleosome-free region in chromatin, using a chromatin assembly extract prepared from *Drosophila* embryos<sup>8</sup>, plasmid DNA, and purified transcription factors. As a model system, we analysed the promoter of the *Drosophila hsp70* gene encoding heat-shock protein 70. The promoter contains sites for interaction with HSF, the heat-shock transcription factor<sup>9</sup>, and GAGA, a constitutively expressed transcription factor that binds to GA/CT-rich sites present in many *Drosophila* genes<sup>10,11</sup>, and for TFIID, the TATA-binding general transcription factor complex<sup>12</sup>; it is also bound to an RNA polymerase II molecule that has paused after synthesizing a short transcript<sup>13,14</sup>. GAGA factor, TFIID and RNA polymerase II are active and can associate with heat-shock promoters *in vitro* in the absence of a heat-shock stimulus<sup>15-18</sup>, so they are potential candidates for establishing an accessible promoter complex poised to respond to the binding of the activated, trimeric form of HSF<sup>9</sup>. Here we describe the role of the GAGA factor in altering chromatin structure. We show that the introduction of GAGA protein during or after nucleosome assembly *in vitro* results in a disruption of nucleosome structure at the *hsp70* promoter. The disruption is characterized by hypersensitivity to DNase I digestion and a realignment of adjacent nucleosomes, and is facilitated by the presence of hydrolysable ATP.

## Disruption of chromatin by GAGA factor

We analysed the ability of GAGA transcription factor<sup>10</sup> to disrupt nucleosome organization using affinity-purified, histidine-tagged GAGA protein expressed in insect cells with a baculovirus expression vector. Recombinant GAGA factor (hereafter called GAGA) was introduced at each of three stages of nucleosome assembly on a 6.2-kilobase (kb) plasmid carrying *hsp70*:

at the onset (zero time, 0 h), at a stage when the assembly of regularly spaced nucleosomes is nearly complete (2.5 h), or when nucleosome assembly has reached a maximum (5.5 h) (ref. 8, and our unpublished observations). The reconstituted plasmid chromatin was tested for the presence of nucleosome organization at specific locations by prolonged digestion with micrococcal nuclease (MNase), followed by gel electrophoresis and blot hybridization with unique oligonucleotide probes. As micrococcal nuclease initially cleaves within the linker DNA between nucleosome core particles and then progressively trims to the core from each end of the nucleosome<sup>19</sup>, the presence of a nucleosome core particle can be gauged by the accumulation of the canonical, 146-base-pair (bp) nuclease-resistant fragment surviving near the limit of digestion.

When the *hsp70*-plasmid DNA was reconstituted in the chromatin assembly reaction without GAGA, the micrococcal nuclease digestion pattern of the promoter region revealed by hybridization (Fig. 1a) clearly shows that an intact nucleosome has been assembled at sequences corresponding to the oligonucleotide probe (the probe is specific for DNA between positions 115 to 132, which partially overlaps two of four GA/CT elements on the *hsp70* promoter) (see also Fig. 1d, probe C). In addition to the 146-bp fragment derived from the nucleosome core particle, a ladder of discrete fragments corresponding to nucleosome oligomers can be seen at intermediate stages of digestion by micrococcal nuclease. This pattern of cleavage indicates that the DNA surrounding the *hsp70* promoter is organized in a regularly spaced (but not necessarily positioned) array of nucleosomes with a characteristic repeat length of ~180 bp<sup>8</sup>.

A dramatically different cleavage pattern at the *hsp70* promoter was seen when GAGA was added at the onset of nucleosome assembly (0 h) in (Fig. 1a). Upon extensive digestion with micrococcal nuclease, the abundance of the 146-bp fragment is decreased up to 5-fold. In addition, subnucleosomal fragments shorter than 146 bp are evident, despite the difficulty in effecting quantitative Southern transfer of very small DNAs. The generation of these subfragments and the loss of the 146-bp fragment represents an invasion and cleavage of the DNA within the nucleosome core particle by micrococcal nuclease, showing that nucleosome organization has been disrupted at the promoter region and giving rise to a smear of DNA sizes at intermediate stages of digestion (Fig. 1a), rather than the repeat pattern of nucleosome oligomers. This ability of the nuclease to invade the disrupted region is probably a result of the secondary trimming action of the enzyme rather than of primary endonucleolytic cleavage (see later).

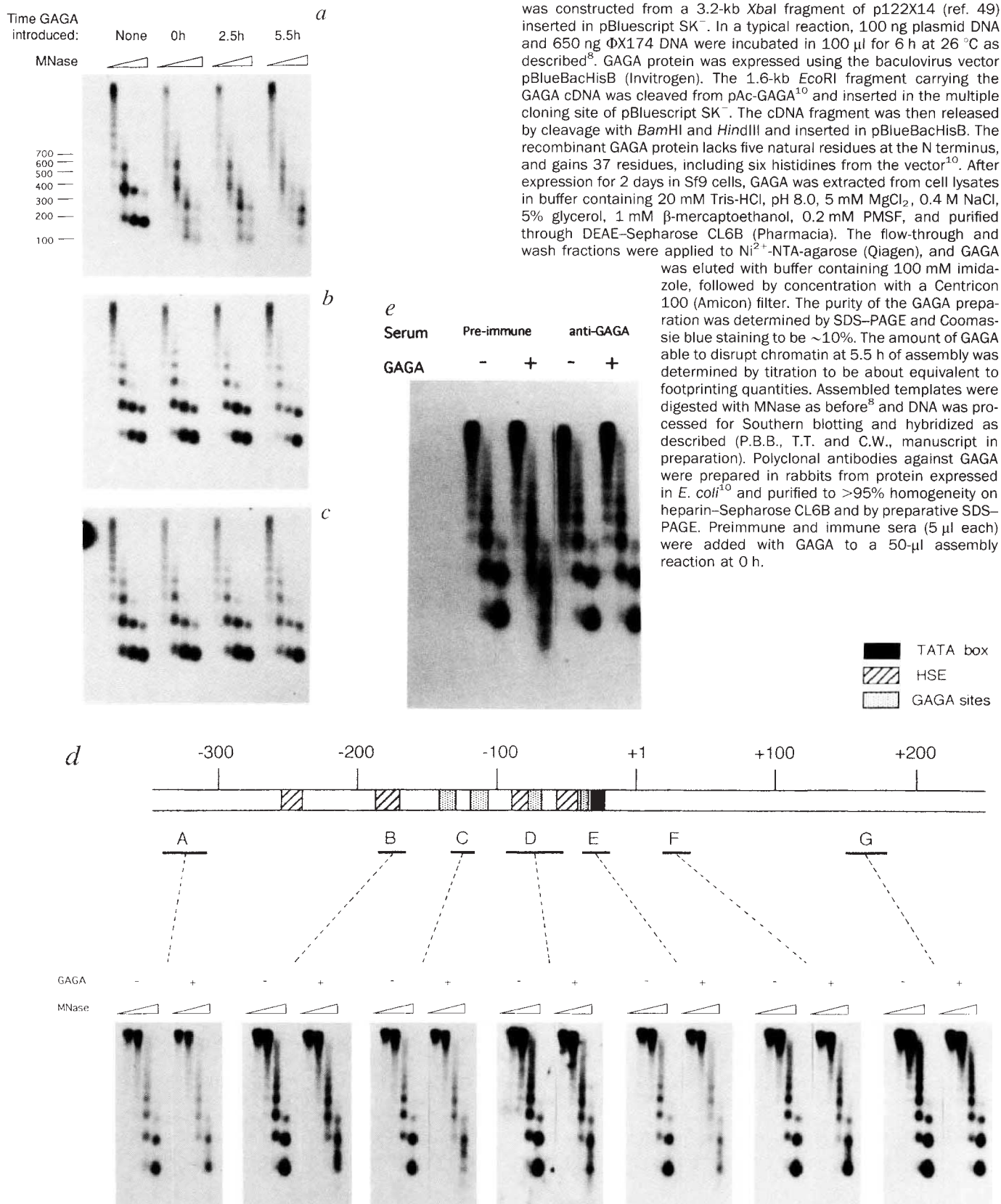
Using these criteria for nucleosome disruption, chromatin

\* Present address: Gene Expression Programme, EMBL, Meyerhofstrasse 1, D-69117 Heidelberg, Germany.

† To whom correspondence should be addressed.



FIG. 1 GAGA-dependent chromatin disruption *in vitro*. a–c, Micrococcal nuclease (MNase) digestion patterns of *hsp70*-plasmid chromatin reconstituted with GAGA as indicated. DNA blots were hybridized sequentially with oligonucleotides: a, (–115 to –132); b, (+1,803 to +1,832); and c, (2,499 to 2,528 of pBluescript SK<sup>–</sup>), respectively. Numbers to the left indicate size calibration (in base pairs). d, MNase digestion patterns of the *hsp70* promoter region reconstituted with GAGA



added at 2.5 h of nucleosome assembly. The blot was sequentially hybridized with the probes A–G (map positions: A, –340 to –311; B, –184 to –165, –132 to –115, D, –89 to –50, E, –36 to –17; F, +19 to +36; G, +148 to +176). Similar results were obtained when GAGA was added at 0 h or after 5.5 h of assembly. e, Antibody inhibition of GAGA function.

**METHODS.** Chromatin assembly extracts were prepared from preblastoderm stage embryos as described<sup>8</sup>. The *hsp70*-plasmid (pdhspXX3.2) was constructed from a 3.2-kb *Xba*I fragment of p122X14 (ref. 49) inserted in pBluescript SK<sup>–</sup>. In a typical reaction, 100 ng plasmid DNA and 650 ng  $\Phi$ X174 DNA were incubated in 100  $\mu$ l for 6 h at 26 °C as described<sup>8</sup>. GAGA protein was expressed using the baculovirus vector pBlueBachHisB (Invitrogen). The 1.6-kb *Eco*RI fragment carrying the GAGA cDNA was cleaved from pAc-GAGA<sup>10</sup> and inserted in the multiple cloning site of pBluescript SK<sup>–</sup>. The cDNA fragment was then released by cleavage with *Bam*HI and *Hind*III and inserted in pBlueBachHisB. The recombinant GAGA protein lacks five natural residues at the N terminus, and gains 37 residues, including six histidines from the vector<sup>10</sup>. After expression for 2 days in Sf9 cells, GAGA was extracted from cell lysates in buffer containing 20 mM Tris-HCl, pH 8.0, 5 mM MgCl<sub>2</sub>, 0.4 M NaCl, 5% glycerol, 1 mM  $\beta$ -mercaptoethanol, 0.2 mM PMSF, and purified through DEAE-Sephacrose CL6B (Pharmacia). The flow-through and wash fractions were applied to Ni<sup>2+</sup>-NTA-agarose (Qiagen), and GAGA was eluted with buffer containing 100 mM imidazole, followed by concentration with a Centricon 100 (Amicon) filter. The purity of the GAGA preparation was determined by SDS-PAGE and Coomassie blue staining to be ~10%. The amount of GAGA able to disrupt chromatin at 5.5 h of assembly was determined by titration to be about equivalent to footprinting quantities. Assembled templates were digested with MNase as before<sup>8</sup> and DNA was processed for Southern blotting and hybridized as described (P.B.B., T.T. and C.W., manuscript in preparation). Polyclonal antibodies against GAGA were prepared in rabbits from protein expressed in *E. coli*<sup>10</sup> and purified to >95% homogeneity on heparin-Sephacrose CL6B and by preparative SDS-PAGE. Preimmune and immune sera (5  $\mu$ l each) were added with GAGA to a 50- $\mu$ l assembly reaction at 0 h.

structure was similarly altered when GAGA was introduced after 2.5 h of nucleosome assembly, and was also significantly affected when GAGA was introduced after the completion of nucleosome assembly (at 5.5 h; Fig. 1a). We conclude that GAGA can alter the structure of nucleosomes whether by direct competition with

the process of nucleosome assembly *in vitro*, or by affecting the structure of nucleosomes previously deposited on DNA. To investigate the effects of GAGA on chromatin structure at locations distant from the *hsp70* promoter, the same DNA blot was stripped of the probe and rehybridized with oligonucleotides

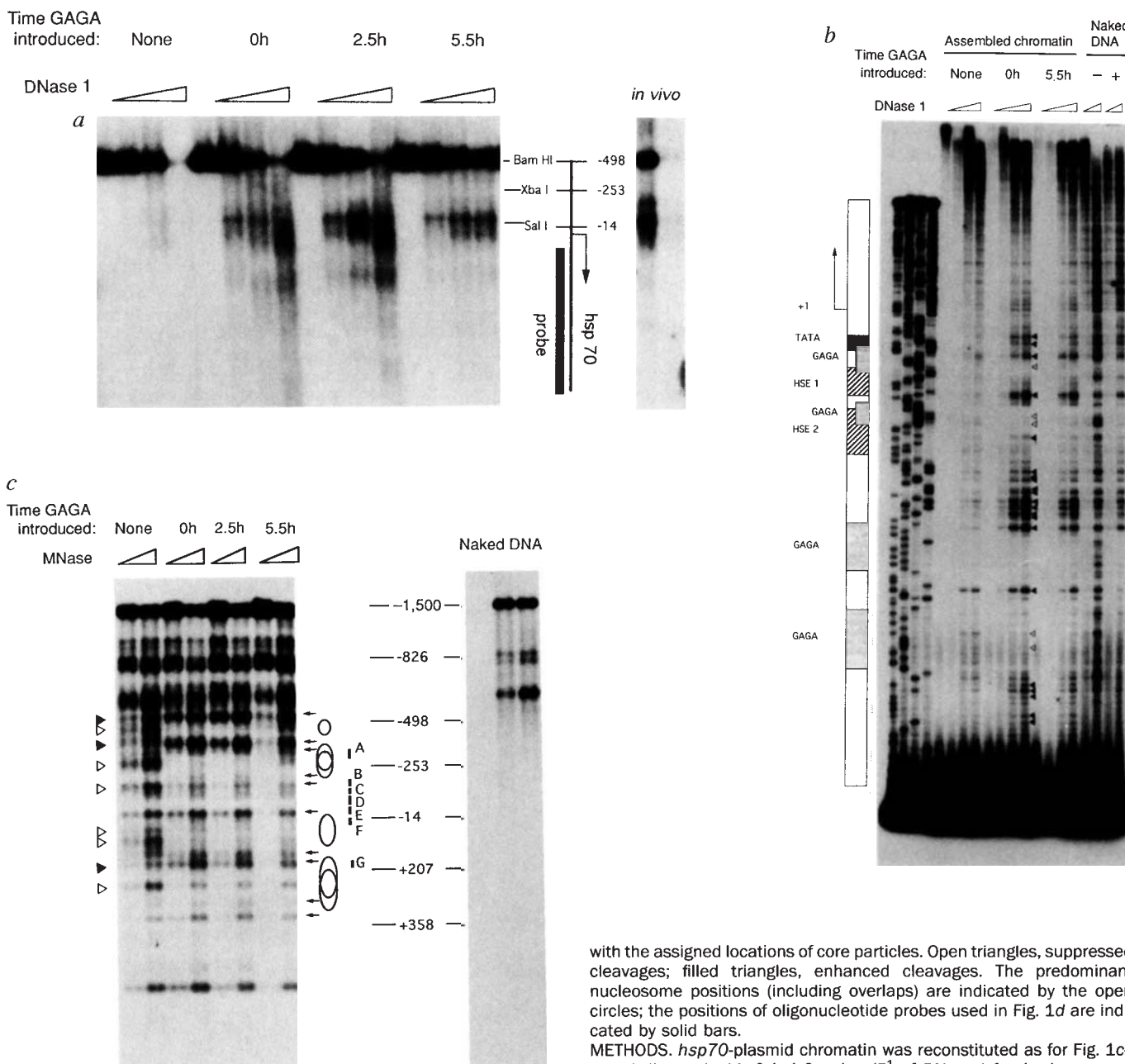


FIG. 2 Mapping DNase I hypersensitivity and nucleosome positions on the reconstituted *hsp70* promoter. *a*, Indirect end-labelling of DNase I cleavages relative to a BamHI site at +1,258 (ref. 20). The location of the probe on a restriction map of *hsp70* is indicated by the solid bar. The *in vivo* sample shows the cleavage pattern of the endogenous *hsp70* genes in 0–24 h embryo nuclei. *b*, Primer extension–linear amplification analysis of DNase I cleavages on a 6% sequencing gel. Hypersensitive nucleotides are indicated by filled triangles, and protected nucleotides by open triangles. The GAGA binding sites, heat-shock elements (HSE) and TATA box are indicated by the stippled, hatched, and solid bars. The first four lanes on the left are sequencing reactions: T, C, G, A. *c*, Indirect end-labelling of MNase cleavages relative to a PstI restriction site at +1,187. Arrows connect MNase cleavages adjacent to the promoter region (minor cleavages are not included)

with the assigned locations of core particles. Open triangles, suppressed cleavages; filled triangles, enhanced cleavages. The predominant nucleosome positions (including overlaps) are indicated by the open circles; the positions of oligonucleotide probes used in Fig. 1d are indicated by solid bars.

**METHODS.** *hsp70*-plasmid chromatin was reconstituted as for Fig. 1c–e and digested with 0.1–1.0 units  $\mu\text{l}^{-1}$  of DNase I for 1 min at room temperature. Purified DNA was digested with BamHI, which cuts at +1,258 and at –498, separated on a 1.2% agarose gel, and transferred to a GeneScreen membrane. The blot was probed with a  $^{32}\text{P}$ -labelled *Nru*I (+32) to *Pst*I (+1,187) fragment. For linear amplification, a  $^{32}\text{P}$ -labelled primer (–184 to –165) was extended with *Taq* polymerase for 30 cycles using the Cycle Sequencing Kit (Perkin Elmer). The reaction was supplemented with dNTPs (10  $\mu\text{M}$  each) in 5  $\mu\text{l}$  (final) according to the manufacturer's directions. For naked DNA controls, the same amount of GAGA was incubated with plasmid DNA before nuclease digestion. For mapping MNase cleavages by indirect end-labelling, DNA from the MNase digestion samples of Fig. 1d were cleaved with *Pst*I (+1,187) and *Sac*II (which cuts in the multiple cloning site at nucleotide 751 of pBluescript, just upstream of the *Xba*I site at –1,515). DNA was separated on a 1.1% agarose gel (24 h electrophoresis in TBE buffer at 1.5V per cm) and the blot was hybridized with an *Nru*I–*Pst*I fragment.



corresponding to the 3' end of the *hsp70* gene and to the ampicillin-resistance gene of the plasmid vector (Fig. 1b and c). GAGA had no effect on micrococcal nuclease digestion patterns in either case, demonstrating the specificity of GAGA-mediated nucleosome disruption; these controls were used as an internal standard for quantifying changes in the micrococcal nuclease digestion pattern at the *hsp70* promoter region.

We next determined the extent of specific disruption along *hsp70* promoter sequences by sequential blot hybridization with oligonucleotide probes spanning the entire promoter from positions -340 to +176 (Fig. 1d; probes A-G). Among these, probes B-E, which cover the upstream region from -184 to -17 (including the four GA/CT elements), revealed significant nucleosome disruption as evidenced by GAGA-dependent loss of the nucleosome monomer fragment upon extensive digestion, with generation of subfragments and a smearing of the nucleosome oligomer ladder. The extent of DNA spanned by these probes (~160 bp) indicates that histone-DNA interactions in the nucleosome core are disturbed over this region, presumably as a result of GAGA binding. We observed a smaller effect of GAGA on the digestion pattern for probes A, F and G; the nucleosomes on these regions are rearranged at a number of restricted locations (see below). To confirm that the disruption of nucleosomes on the *hsp70* promoter was dependent on GAGA and not on other proteins present in the GAGA preparation, we showed that disruption was abolished by the presence of polyclonal antibodies raised against purified bacterially expressed GAGA (Fig. 1e).

## Reconstitution of DNase I hypersensitivity

As revealed by partial DNase I digestion and indirect end-labelling, a broad hypersensitive site with a major peak at about position -100 is reconstituted on the *hsp70* promoter when GAGA is introduced in the assembly reaction (Fig. 2a). DNase I hypersensitivity is observed when GAGA was added at the start (0 h), at 2.5 h, and at the completion of nucleosome assembly (5.5 h). Although the major hypersensitive peak at about position -100 is close to or coincident with the natural peak of hypersensitivity at around position -93 previously mapped on the endogenous *hsp70* genes<sup>20</sup> (Fig. 2a: 'in vivo'), some differences in the fine structure and span of the hypersensitive region can be seen. As the basal transcription factors and RNA polymerase II are deficient in the chromatin assembly extract (unpublished observations), the inclusion of these components may be necessary to reconstitute the whole hypersensitive structure faithfully.

The positions of DNase I cleavage were mapped to single-nucleotide resolution by primer extension-linear amplification analysis (Fig. 2b). Protection from DNase I cleavage was moderate over the GA/CT repeats when GAGA was introduced in the nucleosome assembly reaction, but the sequences between and flanking the GA/CT repeats were strongly hypersensitive (Fig. 2b). The TATA box and two heat-shock control elements are included in the sequences hypersensitive to DNase I. Positions of hypersensitive cleavage are consistent with the low-resolution map of the DNase I-hypersensitive site obtained by indirect end-labelling, and with the region of nucleosome disruption

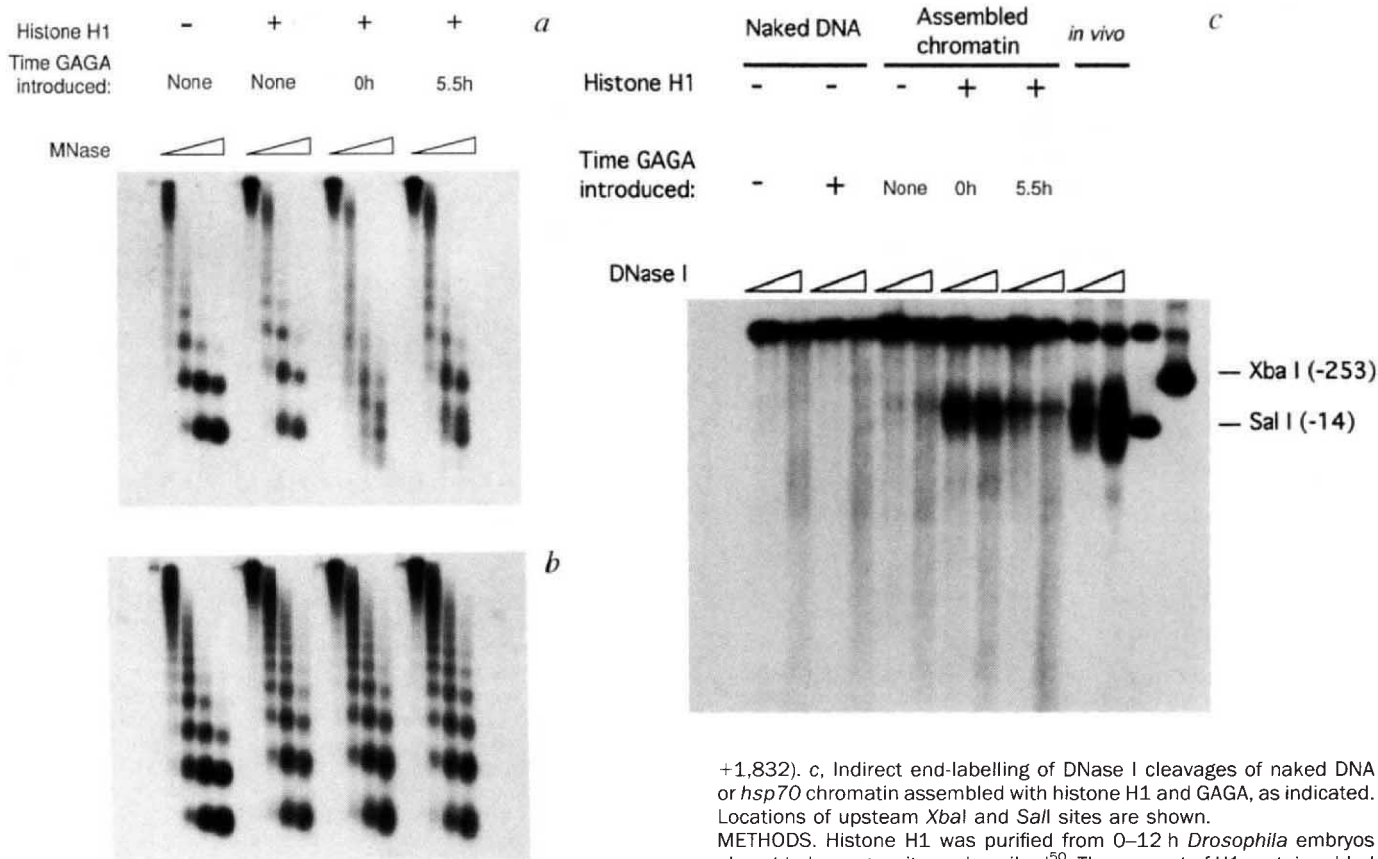


FIG. 3 Effect of GAGA on histone H1-containing chromatin. a, b, MNase digestion patterns of *hsp70* plasmid chromatin reconstituted with GAGA and histone H1 as indicated. DNA was blotted and hybridized sequentially with oligonucleotides: a, (-115 to -132), and b, (+1,803 to

+1,832). c, Indirect end-labelling of DNase I cleavages of naked DNA or *hsp70* chromatin assembled with histone H1 and GAGA, as indicated. Locations of upstream Xba I and Sal I sites are shown.

METHODS. Histone H1 was purified from 0–12 h *Drosophila* embryos almost to homogeneity as described<sup>50</sup>. The amount of H1 protein added to the assembly reaction was titrated so that the average nucleosome repeat was increased to ~195 bp; the amount of H1 incorporated, as estimated by silver staining, is roughly stoichiometric<sup>8</sup>. The reconstitution assay and subsequent analysis are described in the legends to Figs 1 and 2.

tion determined by extensive digestion with micrococcal nuclease.

### Rearrangement of adjacent nucleosomes

We determined the positions of nucleosomes assembled in the vicinity of the *hsp70* promoter by mapping the nucleosome linker regions after partial digestion with micrococcal nuclease digestion and indirect end-labelling (Fig. 2c). When the *hsp70*-carrying plasmid was reconstituted in the absence of GAGA, the initial nuclease cleavages in the vicinity of the *hsp70* promoter were spaced irregularly, reflecting substantial heterogeneity in the positions of the assembled nucleosomes. Inclusion of GAGA at the start, at 2.5 h, or at the completion of nucleosome assembly resulted in suppression and enhancement of cleavages in the regions flanking the site of nucleosome disruption. Changes in the initial cleavage pattern after GAGA binding allow new assignments for the predominant nucleosome positions surrounding the *hsp70* promoter. Hence, as well as disrupting nucleosome organization over its cognate sites, binding of GAGA causes realignment of surrounding nucleosomes, probably by restricting adjacent nucleosomes to a subset of positions. The micrococcal nuclease digestion patterns for oligonucleotide probes A, F and G (Fig. 1d) are consistent with their locations in relation to the rearranged nucleosome core and linker positions. As already mentioned, initial cleavages by micrococcal nuclease mapped by indirect end-labelling do not reveal nucleosome disruption at the promoter as hypersensitivity to the nuclease; in fact, the initial cleavage pattern demonstrates some protection of the promoter region upon GAGA binding.

### Effect on histone H1-containing chromatin

Nucleosomes assembled with the preblastoderm *Drosophila* embryo extract are deficient in the major linker histone H1, which is apparently synthesized only during post blastoderm development<sup>21</sup>. We have shown previously that purified histone H1 can be incorporated during nucleosome assembly, leading to an increased repeat length of ~197 bp<sup>8</sup>. Under these assembly conditions, GAGA is effective in specifically disrupting nucleosome structure on the *hsp70* promoter when introduced at the onset of assembly (Fig. 3a, b). But when GAGA is added at the completion of assembly (5.5 h), disruption is less (Fig. 3a). Disruption of H1-containing chromatin was accompanied by the formation of a DNase I-hypersensitive site, which was less prominent when GAGA was introduced at the completion of assembly (Fig. 3c). Thus the ability of GAGA to dislodge pre-assembled nucleosomes at the *hsp70* promoter is attenuated in chromatin containing histone H1.

### Promoter requirements for disruption

We tested the sequence requirements for nucleosome disruption by reconstituting plasmids carrying deletions of the *hsp70* upstream region (Fig. 4a). The effect of GAGA on plasmid *pdhsp*Δ186, which includes the four GA/CT elements within sequences -186 to +296 of the *hsp70* gene, was essentially the same as the effect on the original *hsp70* plasmid containing the entire coding region and 1.5 kb of upstream DNA (Fig. 4b). But when the template carried a 5' deletion to -90, removing the two distal GA/CT elements (*pdhsp*Δ90), there was disruption when GAGA was added at the start but not at the completion

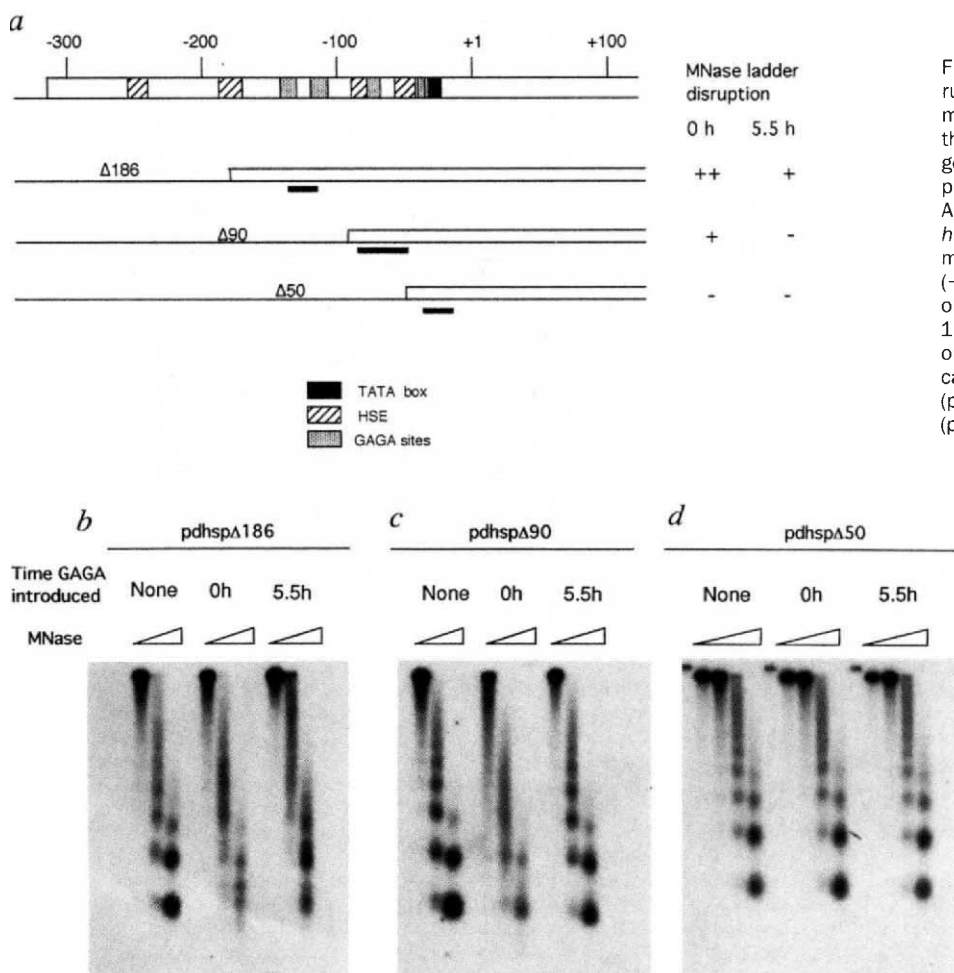


FIG. 4 Promoter elements for chromatin disruption. **a**, Restriction map showing *hsp70* promoter deletions. Vector DNA is represented by the thin line, and open bars show the *hsp70* gene. Oligonucleotide probes used for each plasmid template are indicated by solid bars. All constructs contain up to +296 bp of the *hsp70* sequence. The relative extent of chromatin disruption is summarized on the right; (++) is the designated level of disruption observed on plasmid *dhspXX3.2*, which carries 1.5 kb of upstream DNA. **b–d**, Effects of GAGA on the assembly of *hsp70* promoter constructs carrying upstream sequences to -186 (*pdhsp*Δ186), -90 (*pdhsp*Δ90) and -50 (*pdhsp*Δ50), respectively. Plasmid DNAs were reconstituted with GAGA, digested with MNase, and DNA blots were hybridized with oligonucleotides C, D and E as for Fig. 1.

**METHODS.** Assembly reactions and subsequent analyses are described in Fig. 1 legend, except that 50 ng promoter plasmid DNA and 700 ng ΦX174 DNA were used.



of assembly (Fig. 4c). When the *hsp70* promoter was deleted to -50, which removes the third GA/CT element (pdhsp $\Delta$ 50), no disruption by GAGA was discernible (Fig. 4d). Therefore promoter sequences that include at least two GA/CT elements are necessary for nucleosome displacement when GAGA is competing directly with the nucleosome assembly reaction, and sequences including three to four elements are required for the disruption of a preassembled nucleosome.

### ATP requirement

The assembly of nucleosomes *in vitro* requires ATP and an energy-regeneration system<sup>8</sup>. We investigated the energy require-

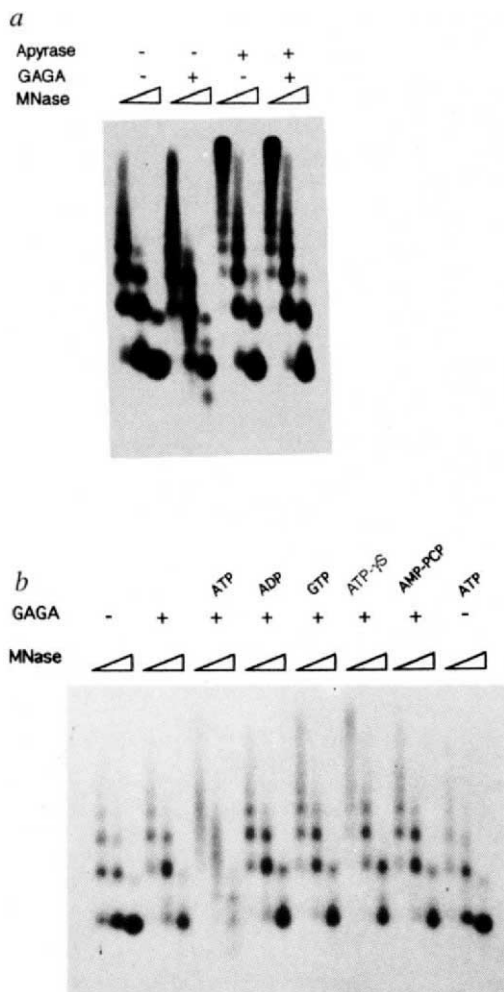


FIG. 5 ATP requirement for GAGA-dependent disruption of chromatin structure. **a**, Effect of apyrase on chromatin disruption at the *hsp70* promoter; *hsp70*-plasmid chromatin reconstituted with GAGA and treated with apyrase was analysed by MNase digestion. The resultant DNA was blotted and hybridized as for Fig. 1a. **b**, Restoration of nucleosome disruption with supplemental ATP. Partially purified *hsp70* chromatin was incubated with GAGA, nucleotides and analogues as indicated, digested with MNase, and processed as described.

**METHODS.** After 5.5 h of reconstitution of the *hsp70*-plasmid in 50  $\mu$ l of chromatin assembly reaction, 0.1 units of apyrase (Sigma, grade VI) was added and incubated at 26 °C for 15 min before incubation with GAGA for 30 min, followed by MNase digestion (Fig. 1 legend). For purification of chromatin, 100- $\mu$ l aliquots of a 6-h assembly were applied on 1-ml Bio-Gel A-1.5 m spin columns (prepared in a 1-ml tuberculin syringe) pre-equilibrated with 10 mM HEPES-KOH, pH 7.7, 50 mM KCl, 0.5 mM EGTA, 5 mM  $MgCl_2$ , 10% glycerol. 50  $\mu$ l of the fractionated material was incubated at 26 °C for 30 min with GAGA and nucleotides (0.8 mM) as indicated.

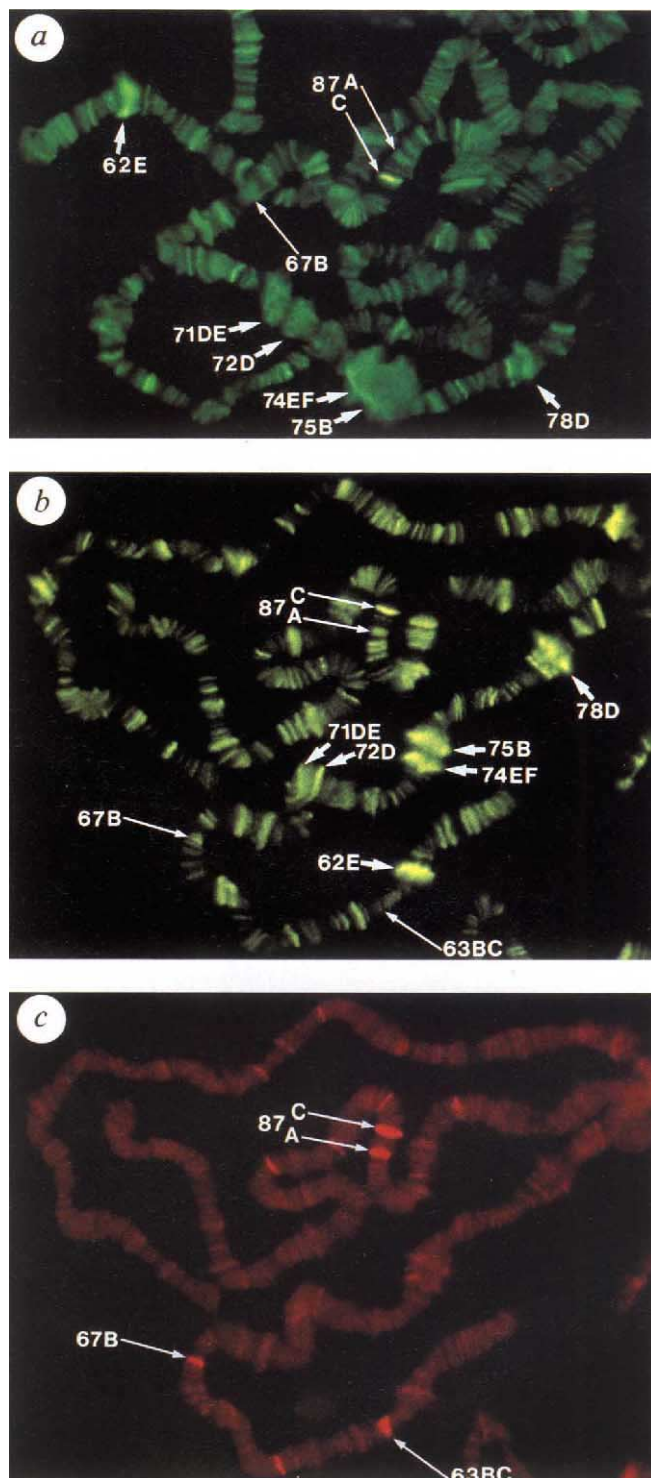


FIG. 6 Localization of GAGA on polytene chromosomes. **a**, **b**, Indirect immunofluorescent staining for GAGA (fluorescein isothiocyanate, FITC) on polytene chromosomes prepared from **a**, unshocked, and **b**, heat-shocked (2 min at 37 °C) third instar larvae. Loci are indicated that carry *hsp70* genes (87A, 87C), the *hsp82* gene (63BC), the small heat-shock genes *hsp27*, *hsp26*, *hsp23* and *hsp22* (67B), prominent developmental puffs. **c**, The same preparation as in **b**, stained for HSF with rhodamine.

**METHODS.** Polytene squashes and chromosome staining were performed as described<sup>51</sup>, using an additional pre-fixing step. Dry milk (Carnation) was substituted for BSA. A 1:500 dilution of rabbit anti-serum to GAGA, and a 1:1,000 dilution of mouse antiserum to HSF were used as primary antibodies.

ments of chromatin disruption by adding the ATP-hydrolysing enzyme apyrase after completion of nucleosome assembly but before introduction of GAGA. Treatment of apyrase substantially suppressed nucleosome disruption by GAGA on the *hsp70* promoter (Fig. 5a); treatment of GAGA alone with apyrase did not affect its ability to bind to free DNA in an electrophoretic mobility shift assay (data not shown). Depletion of ATP using hexokinase and glucose also suppressed nucleosome disruption (data not shown). GAGA-dependent nucleosome disruption was restored upon addition of fresh ATP to a reconstituted chromatin template depleted of nucleotides by gel filtration (Fig. 5b). The ATP in the reaction could not be substituted with GTP, ADP, or with the non-hydrolysable analogues ATP- $\gamma$ S and AMP-PCP. We conclude that the specific disruption of nucleosome structure by GAGA is facilitated by ATP hydrolysis.

### Chromosomal distribution of GAGA

In addition to the *hsp70* and *hsp26* genes, GAGA binds *in vitro* to a range of housekeeping and developmental genes in *Drosophila*<sup>10</sup>. We determined the distribution for GAGA *in situ* by indirect immunofluorescent staining of polytene chromosomes using a polyclonal antiserum specific for GAGA. The distribution of GAGA was essentially identical between chromosome preparations from unshocked or briefly heat-shocked larvae (Fig. 6a, b). Staining for GAGA was strong at many chromosomal loci, including the 87C locus, which carries several *hsp70* genes. Staining was moderate at 87A, which carries two *hsp70* genes, and at 67B, the site of the small heat-shock genes *hsp27*, *hsp26*, *hsp23* and *hsp22*. Not surprisingly, the staining for GAGA at locus 63BC, which encodes Hsp82 and whose promoter region is deficient in GA/CT repeats<sup>22</sup>, is low to undetectable; other *trans*-acting factors must therefore be responsible for the three DNase I hypersensitive sites and an array of nucleosomes positioned on this heat-shock gene<sup>4</sup>. The co-localization of GAGA and HSF at 87A/87C and 67B but not at 63BC, was confirmed by staining the chromosome preparation for both factors (Fig. 6b, c). Staining for GAGA was very strong at prominent developmental puffs active in the late larval stage: 62E, 71DE, 72D, 74EF, 75B and 78D<sup>23</sup>. Among these loci is the *E74* gene at 74EF, encoding an ETS-related DNA-binding protein that carries multiple sites for GAGA binding *in vitro*<sup>24</sup>.

### Discussion

We have shown that the introduction of recombinant GAGA protein during or after the assembly of long nucleosome arrays *in vitro* leads to a local disruption of chromatin structure at the *hsp70* promoter. The binding of GAGA to four sites on this promoter results in nucleosome disruption, the acquisition of DNase I hypersensitivity at the TATA box and the heat-shock control elements, and a rearrangement of adjacent nucleosomes. Considered together with the chromosomal localization of GAGA *in vivo* at several heat-shock loci under non-stress and heat-stress conditions, our results indicate that this constitutive transcription factor plays a key role in forming an *hsp70* promoter structure that is accessible to the basal transcription factors and activated HSF trimers. The widespread chromosomal distribution of GAGA may reflect a similar functional requirement for other genes in *Drosophila*. Our findings are consistent with results showing that *Drosophila* transformants with mutations in upstream GA/CT elements have decreased nuclease hypersensitivity and promoter activity<sup>25</sup>, and with the functioning of GAGA in an *in vitro* transcription assay<sup>26</sup> by antirepression rather than direct activation.

How does GAGA mediate its function in chromatin? The disruption of nucleosome structure resulting from GAGA binding can occur in competition with the assembly of nucleosomes as well as on pre-existing nucleosomes. In this respect, disruption can be classified as 'dynamic', in contrast to the 'pre-emptive' mechanisms ascribed previously to factor-dependent chromatin perturbation *in vitro*<sup>1</sup>. Dynamic mechanisms of

nucleosome disruption could include dissociation of the nucleosome octamer into the (H3/H4)<sub>2</sub> tetramer and H2A/H2B dimers, sliding of intact octamers away from promoter sequences by weakening of histone DNA interactions, or removal of histone octamers from the promoter. Protein-protein interactions between GAGA proteins bound at several adjacent sites may also subject the intervening DNA to torsional constraints and contribute to nucleosome destabilization. The 519-residue GAGA open reading frame<sup>10</sup> shows a single zinc-finger in the C-terminal region, stretches of glutamine residues and basic amino acids, and an N-terminal 120-residue domain with significant sequence homology to *Drosophila trans*-acting factors *tramtrack*<sup>27–29</sup> and the *Broad Complex*<sup>30</sup>, and *kelch*<sup>31</sup>, a component of intercellular bridges in *Drosophila*. It will be interesting to define the structural domains of GAGA responsible for the effects on chromatin.

The dependence on ATP of GAGA-mediated nucleosome disruption suggests a number of energy-dependent mechanisms to alter histone-DNA contacts. GAGA might itself bind ATP and disrupt nucleosome structure by a conformational change; but it contains no canonical ATP-binding motif, neither have we been able to demonstrate ATP-binding activity for this protein. GAGA may act in conjunction with other components in the crude embryo extract that need to hydrolyse ATP for nucleosome disruption. The evolutionarily conserved proteins of the SWI 2/SNF 2 family are non-DNA-binding mediators that facilitate transcriptional activation by antagonizing the inhibitory effects of chromatin proteins<sup>32,33</sup>. As these proteins share conserved motifs with poxvirus DNA-dependent ATPases<sup>34</sup>, homologous *Drosophila* proteins such as *brahma*<sup>35</sup> might assist GAGA by fulfilling the ATP-dependent function. Alternatively, other ATP-dependent factors or enzymes affecting nucleosome assembly or spacing (for review, see ref. 36) may act constitutively or be locally concentrated by GAGA to modify or destabilize the nucleosome core particle such that binding of the transcription factor at multiple sites would suffice *per se* to complete disruption. In this respect, we have noted minor but reproducible effects on nucleosome structure mediated by GAGA binding in extracts depleted of ATP (Fig. 5, and unpublished observations). Before individual components and their relative contributions to nucleosome disruption can be defined and different models distinguished, fractionated material and a defined chromatin substrate will be needed.

Although nucleosome disruption by dynamic competition *in vivo* is feasible for several yeast *trans*-acting factors (PHO 4 (ref. 37), GRF-2/REF-1 (refs 38, 39), RAP-1 (ref. 40) and GAL4 (ref. 41)), the means by which disruption can occur has only been investigated *in vitro* for GAL4, for which nucleosome displacement required the presence of excess carrier DNA and was independent of the GAL4 transactivation domain<sup>42</sup>. In addition, a GAL4-VP16 hybrid protein can relieve nucleosome-mediated repression of transcription *in vitro*<sup>43–45</sup>. The binding of the glucocorticoid receptor in responsive cells has also been implicated in the active process of nucleosome disruption, although this receptor forms a stable ternary complex with nucleosomal DNA *in vitro*<sup>46–48</sup>. Our finding that GAGA-mediated disruption of existing nucleosomes is facilitated by ATP hydrolysis opens avenues of investigation into how the repressive effects of nucleosome structure at a eukaryotic promoter might be overcome. □

Received 13 September; accepted 22 December 1993.

1. Felsenfeld, G. *Nature* **335**, 219–224 (1992).
2. Grunstein, M. *Rev. Cell. Biol.* **6**, 643–678 (1990).
3. Kornberg, R. D. & Lorch, Y. A. *Rev. Cell. Biol.* **8**, 563–587 (1992).
4. Wu, C. *Nature* **286**, 854–860 (1980).
5. Gross, D. S. & Garrard, W. T. A. *Rev. Biochem.* **57**, 159–197 (1988).
6. Workman, J. L. & Buchman, A. R. *Trends biochem. Sci.* **18**, 90–95 (1993).
7. Hayes, J. J. & Wolffe, A. P. *Bioessays* **14**, 597–603 (1992).
8. Becker, P. B. & Wu, C. *Molec. cell. Biol.* **12**, 2241–2249 (1992).
9. Lis, J. T. & Wu, C. *Cell* **74**, 1–4 (1993).
10. Soeller, W. & Kornberg, T. *Molec. cell. Biol.* **13**, 7961–7970 (1993).
11. Biggin, M. & Tjian, R. *Cell* **53**, 699–711 (1988).
12. Dynlacht, B. D., Hoey, T. & Tjian, R. T. *Cell* **66**, 563–576 (1991).



13. Rougvié, A. E. & Lis, J. T. *Cell* **54**, 795–804 (1988).
14. Rasmussen, E. & Lis, J. T. *Proc. natn. Acad. Sci. U.S.A.* **90**, 7923–7927 (1993).
15. Gilmour, D. S., Thomas, G. H. & Elgin, S. C. R. *Science* **245**, 1487–1490 (1989).
16. Wu, C. *Nature* **317**, 84–87 (1985).
17. Purnell, B. A. & Gilmour, D. S. *Molec. cell. Biol.* **13**, 2593–2603 (1993).
18. Parker, C. S. & Topol, J. *Cell* **36**, 357–369 (1984).
19. van Holde, K. E. *Chromatin* 16–30 (Springer, New York, 1989).
20. Wu, C. *Nature* **309**, 229–234 (1984).
21. Elgin, S. C. R. & Hood, L. E. *J. biol. Chem.* **12**, 4984–4991 (1973).
22. Blackman, R. K. & Meselson, M. *J. molec. Biol.* **188**, 499–515 (1986).
23. Ashburner, M. in *Developmental Studies on Giant Chromosomes* (Springer, Berlin, 1972).
24. Burtis, K. C. et al. *Cell* **61**, 85–89 (1990).
25. Lue, Q., Wallrath, L. L., Granok, H. & Elgin, S. C. R. *Molec. cell. Biol.* **13**, 2802–2814 (1993).
26. Kerrigan, L. A., Croston, G. E., Lira, L. & Kadonaga, J. T. *J. biol. Chem.* **266**, 574–582 (1991).
27. Harrison, S. D. & Travers, A. A. *EMBO J.* **9**, 207–216 (1990).
28. Read, D. & Manley, J. L. *EMBO J.* **11**, 1035–1044 (1992).
29. Brown, J. L. & Wu, C. *Development* **117**, 45–58 (1993).
30. DiBello, P. R. et al. *Genetics* **129**, 385–397 (1991).
31. Xue, F. & Cooley, L. *Cell* **72**, 681–693 (1993).
32. Peterson, C. L. & Herskowitz, I. *Cell* **68**, 573–583 (1992).
33. Winston, F. & Carlson, M. *Trends Genet.* **8**, 387–391 (1992).
34. Henikoff, S. *Trends biochem. Sci.* **18**, 291–292 (1993).
35. Tamkun, J. W. et al. *Cell* **68**, 561–572 (1992).
36. Aimouzni G. & Wolffe, A. *Expl Cell Res.* **205**, 1–15 (1993).
37. Schmid, A., Fascher, K. D. & Horz W. *Cell* **71**, 1–20 (1992).
38. Chasman, D. I. et al. *Genes Dev.* **4**, 503–514 (1990).
39. Morrow, B. E., Ju, Q. & Warner, J. R. *Molec. cell. Biol.* **13**, 1173–1182 (1993).
40. Devlin, C., Tice-Baldwin, K., Shore, D. & Arndt K. T. *Molec. cell. Biol.* **11**, 3642–3651 (1991).
41. Axelrod, J. D., Reagan, M. S. & Majors, J. *Genes Dev.* **7**, 857–869 (1993).
42. Workman, J. L. & Kingston, R. E. *Science* **258**, 1780–1784 (1992).
43. Workman, J. L., Taylor, I. A. C. & Kingston R. E. *Cell* **64**, 533–544 (1991).
44. Kamakaka, R. T., Bulger, M. & Kadonaga, J. T. *Genes Dev.* **7**, 1779–1795 (1993).
45. Lorch, Y., LaPointe, J. W. & Kornberg, R. D. *Genes Dev.* **6**, 2282–2287 (1992).
46. Perlman, T. & Wragne, O. *EMBO J.* **7**, 3073–3083 (1988).
47. Pina, B., Bruggemeier, u. & Beato, M. *Cell* **60**, 719–731 (1990).
48. Archer, T. K. et al. *Molec. cell. Biol.* **11**, 688–698 (1991).
49. Mason, P. J. et al. *J. molec. Biol.* **156**, 21–35 (1982).
50. Croston, G. E., Lira, L. M. & Kadonaga, J. T. *Prot. Expr. Purific.* **2**, 162–169 (1991).
51. Westwood, J. T., Clos, J., & Wu, C. *Nature* **353**, 822–827 (1991).

ACKNOWLEDGEMENTS. We thank W. Soeller and T. Kornberg for the GAGA cDNA clone and for access to data before publication; J. Wisniewski and M. Wisniewska for assistance in expressing GAGA from the baculovirus vector; and members of our laboratory for discussion.

# Insight into E-selectin/ligand interaction from the crystal structure and mutagenesis of the lec/EGF domains

**Bradford J. Graves, Robert L. Crowther, Chitra Chandran, John M. Rumberger\*, Shirley Li, Kuo-Sen Huang, David H. Presky, Philip C. Familletti, Barry A. Wolitzky & Daniel K. Burns\***

Roche Research Center, Hoffmann-La Roche Inc., Nutley, New Jersey 07110, USA

**The three-dimensional structure of the ligand-binding region of human E-selectin has been determined at 2.0 Å resolution. The structure reveals limited contact between the two domains and a coordination of  $\text{Ca}^{2+}$  not predicted from other C-type lectins. Structure/function analysis indicates a defined region and specific amino-acid side chains that may be involved in ligand binding. These features of the E-selectin/ligand interaction have important implications for understanding the recruitment of leukocytes to sites of inflammation.**

E-SELECTIN is a cytokine-inducible, endothelial-cell-specific membrane glycoprotein capable of mediating the adhesion of neutrophils<sup>1</sup>. It is a member of the selectin gene family which includes L-selectin<sup>2–4</sup> and P-selectin<sup>5</sup>. The selectins constitute a highly conserved gene family sharing a common structural organization and function during inflammation. Each has been implicated in the initial 'rolling' step of leukocyte extravasation (reviewed in refs 6–8). The selectins contain an amino-terminal lectin domain (lec), followed by an epidermal growth factor-like element (EGF), a variable number of complement regulatory repeat units (cr), a single membrane-spanning region, and a carboxy-terminal cytoplasmic domain. The lectin domain is analogous to other C-type  $\text{Ca}^{2+}$ -dependent animal lectins, suggesting that selectin-mediated adhesion of leukocytes is a consequence of carbohydrate-protein interactions. Although a number of distinct potential selectin-specific ligands have been described<sup>9–12</sup>, all selectins can bind the carbohydrate structure sialyl Lewis X ( $\text{sLe}^{\text{x}}$ )<sup>13</sup>.

The lectin and EGF domains are both necessary and sufficient to mediate neutrophil adhesion<sup>14–16</sup>. To elucidate the molecular interactions between E-selectin and its ligand, we have solved

the three-dimensional structure of the lec/EGF fragment of E-selectin and evaluated the role of specific amino-acid side chains in neutrophil adhesion.

## Structure determination and description

A soluble form of E-selectin containing the lec/EGF domains was produced in CHO cells, purified as described<sup>17</sup>, and deglycosylated with *N*-glycanase. Deglycosylation of E-selectin does not affect its adhesion activity. Crystals of E-selectin lec/EGF diffract beyond 2.0 Å resolution and yield excellent diffraction data. Details of the crystallization and heavy-atom phasing will be published elsewhere (R.L.C. and B.J.G., manuscript in preparation), but the data collection and X-ray refinement statistics are presented in Table 1. The strong phasing by the heavy-atom derivatives can be demonstrated by the clarity of the original multiple isomorphous replacement (MIR) map (Fig. 1a). Main-chain electron density for the 157 amino acid residues was continuous, with the exception of a minor break after Cys 122. In addition, electron density for most side chains and even some water molecules was readily discernible, simplifying the tracing of the map. The refinement process was straightforward, with the current model containing 157 amino acid residues with good geometry, three calcium ions, two chloride ions, and 112 water molecules (Table 1). Full atomic coordinates are being deposited

\* Present address: Glaxo Research Institute, Department of Cell Physiology, Research Triangle Park, North Carolina 27709, USA.

in the Brookhaven Protein Data Bank.

Ribbon drawings of the backbone of the E-selectin lec/EGF domains (Fig. 1*b, c*) indicate that the overall folding of the lec domain is very similar to the lectin domain of the rat mannose-binding protein (MBP)<sup>18</sup>. Likewise, the EGF domain has the same general fold and arrangement of disulphide bonds as other EGF-like domains<sup>19–25</sup>. One of the most interesting features of the lec/EGF structure is the lack of interaction between the two domains (Fig. 1*c*). Only residues 135–139 of the EGF domain make contact with the lec domain, including several hydrogen bonds and a limited number of van der Waals contacts. Most of these hydrogen bonds involve main-chain atoms—the side chains of Glu 135 and Gln 30 being the lone exceptions. The N-terminal ammonium group is a triple hydrogen-bond donor, with the acidic side chain of Glu 135 being one acceptor. Thus, although the contacts are limited, it is anticipated that the relative orientation of the two domains is not likely to change dramatically.

As a C-type lectin, the nature of the calcium binding is a crucial functional aspect of the lec domain. Under the crystallization conditions used here, the structure of lec/EGF includes three calcium ions—all within the lec domain (Fig. 1*b*). One of the calcium ions is well coordinated by the side chains of Glu 80, Asn 82, Asn 105 and Asp 106, as well as the main-chain carbonyl of Asp 106 and two water molecules, to form a pentagonal bipyramid coordination sphere. This will subsequently be referred to as site 1, or the 'bound calcium' in E-selectin. The other two calcium-binding sites are probably adventitious, resulting from the crystallization conditions, and were not observed in isomorphous crystals grown in the absence of added calcium (data not shown).

### Comparison with MBP and EGF domains

Inspection of the E-selectin structure reveals some important differences in the structures of the lec domain and mannose-binding protein (MBP). The structure of MBP in its complexed form<sup>26</sup> will be used for structural comparison because it is at higher resolution, is more highly refined, and has calcium instead of holmium present.

Although there is only ~30% identity between the two sequences, the disulphide bonds and additional clusters of identical residues enable the lectin domains of E-selectin and MBP to be topologically aligned. Based on this alignment and using all topologically equivalent sites (109 *Ca* pairs), a r.m.s. deviation of 1.94 Å is calculated for the best molecular fit. In the resulting overlay of the *Ca* structures (Fig. 2*a*), it is apparent where the two structures diverge. These structural alterations were to be expected, as the selectins and mannose-binding proteins are only distantly related C-type lectins<sup>27</sup>. Many of the differences are due, in part, to the various insertions and deletions. The first major insertion in the E-selectin sequence occurs at residues 45–47, which increases the loop size between the second  $\alpha$ -helix and the second  $\beta$ -strand. The other major insertion occurs at Ile 95–Lys 99. Not only are there five additional residues in this region, but they adopt an interesting conformation: rather than extending the  $\beta$ -strand structures ( $\beta$ 4 and  $\beta$ 5), the additional sequence is bent so that residues Arg 97, Glu 98 and Lys 99 are extended towards the conserved calcium-binding site. The structure of these insertions is of particular interest because residues that are within or adjacent to these loops (Tyr 48, Tyr 94 and Arg 97) have been implicated in neutrophil adhesion (see later).

The last major region in which there is a large disparity in the position of topologically equivalent residues occurs in the calcium-binding loop for residues 84–87 (Fig. 2*c*). In this case, although there is no insertion or deletion, a calcium-binding site (referred to as site 1 in the MBP structure) has been lost and a pair of glycines in MBP are glutamine and aspartic acid in E-selectin. Both Gly 190 (58°, –131°) and Gly 192 (125°, 143°) of MBP adopt  $\phi$ ,  $\psi$  angles that are energetically unfavourable for

TABLE 1 Data collection and refinement statistics for lec/EGF E-selectin

| Resolution limits (Å) | Independent reflections | Completeness (%) | Fraction >3 $\sigma$ (F) | Refinement <i>R</i> -factor |                        |
|-----------------------|-------------------------|------------------|--------------------------|-----------------------------|------------------------|
|                       |                         |                  |                          | Using all data              | Using $F > 3\sigma(F)$ |
| 10.0–3.93             | 1,823                   | 100.0            | 0.99                     | 0.174                       | 0.173                  |
| 3.93–3.15             | 1,727                   | 100.0            | 0.96                     | 0.144                       | 0.140                  |
| 3.15–2.76             | 1,742                   | 99.9             | 0.93                     | 0.157                       | 0.151                  |
| 2.76–2.51             | 1,704                   | 99.9             | 0.89                     | 0.173                       | 0.164                  |
| 2.51–2.34             | 1,693                   | 99.8             | 0.88                     | 0.177                       | 0.167                  |
| 2.34–2.20             | 1,680                   | 99.9             | 0.84                     | 0.190                       | 0.173                  |
| 2.20–2.09             | 1,688                   | 99.7             | 0.80                     | 0.202                       | 0.182                  |
| 2.09–2.00             | 1,661                   | 99.1             | 0.75                     | 0.208                       | 0.186                  |
| 10.0–2.00             | 13,718                  | 99.8             | 0.88                     | 0.173                       | 0.164                  |

Model geometry: r.m.s. dev. bond dist., 0.010 Å; r.m.s. dev. bond angles, 2.60°

Data collection: Data were collected to slightly better than 2.0 Å resolution at room temperature using twin multiwire area detectors (Area Detector Systems). Data from two crystals were merged to attain a high level of completeness. Individual data sets have  $R_{\text{sym}}$  values of 0.044 and 0.059 and average redundancies of 2.2 (25,713 observations of 11,901 reflections) and 2.8 (40,831 observations of 14,399 reflections). The second data collection was made in the reverse order to achieve additional protection against crystal decay effects. The merging *R*-factor between the two data sets is 0.098. Refinement: The structure was refined by the conjugate gradient minimization method as implemented in XPLOR<sup>45</sup>. About 10% of the data was flagged for the purpose of checking refinement with the free *R* value<sup>46</sup>. The *R*(free) for refinement with no sigma cutoff is 0.249 and 0.239 for data with  $F > 3\sigma(F)$ . The final model comprises all 157 amino-acid residues, 112 water molecules, 3 calcium ions and 2 chloride ions.

non-glycine residues; this was noted even for some model-derived structures of E-selectin<sup>28</sup>. Another marked difference in this region occurs at Asn 83. Even though the *Ca* position of Asn 83 overlays the *Ca* of Asp 188 closely, the  $\phi$ ,  $\psi$  angles of the two residues are quite different. A best-molecular-fit calculation that does not include residues Arg 84–Asp 87 (as well as six other positions from the inserted loops) yields a r.m.s. deviation of 1.33 Å for the remaining 99 *Ca* pairs. Thus although a few limited segments diverge extensively, most of the E-selectin lec domain satisfactorily overlays the lectin domain of MBP.

One of the most striking differences between the lectin domains of E-selectin and MBP is the nature of calcium binding. Comparison of Figs 1*b* and 2*a* of ref. 26 shows that only the calcium at site 1 (site 2 in MBP) is shared. All residues in MBP that are involved in binding at this common site are exactly conserved in E-selectin, but owing to the loss of the other calcium-binding site and the altered conformation of the calcium-binding loop, the molecular interactions are different; that is, all of the coordinating residues in the MBP (Glu 185, Asn 187, Glu 193, Asn 205 and Asp 206) are identical in E-selectin (Glu 80, Asn 82, Glu 88, Asn 105 and Asp 106). However, the position and orientation of Glu 88 are altered such that it cannot be a ligand for this calcium ion (Fig. 2*c*). This coordination site has been replaced by a water molecule which is also hydrogen-bonded to Asn 83. The involvement of Asn 83 is an unexpected feature, as the equivalent of Asn 83 in MBP (Asp 188) coordinates to the other calcium (site 1), one that does not exist in E-selectin. The selectins are probably unique among the known C-type lectins in having only one bound calcium. Alignment of the four classes of C-type lectins indicates that the residues involved in calcium binding at site 1 in MBP are highly conserved except in the selectins<sup>18</sup>.

The overall fold of the E-selectin EGF domain is very similar to that of other EGF or EGF-like domains<sup>19–25</sup>. Although there are no available crystallographic structures for an EGF or EGF-like domain (a crystal structure of a fragment of factor Xa has appeared<sup>29</sup>), there are several NMR-determined structures. If



the six cysteine  $\alpha$  positions are used to overlay the structures (Fig. 3a), r.m.s. fits for the six positions range from 0.33 to 1.07 Å for E-selectin versus factor IX (ref. 24), factor X (ref. 25), human transforming growth factor- $\alpha$  (TGF- $\alpha$ ; ref. 23) and murine EGF (ref. 20). The disulphide bonds clearly constrain these domains into compact structures, with significant variability in the loops between cysteines. The similarity to the human clotting factor IX EGF domain (Fig. 3b) is particularly interesting because of the ability of this domain to substitute for the E-selectin EGF domain while maintaining neutrophil adhesion (D.K.B. and B.A.W., manuscript in preparation).

### Structure/activity relationship of E-selectin

To approximate the natural adhesion components, the effect of various mutations on human neutrophil adhesion was evaluated with membrane-associated E-selectin mutants. Only those mutations that did not significantly alter the structure, as judged by monoclonal antibody binding, will be discussed. The results of our membrane-associated E-selectin mutagenesis (Table 2) are in general agreement with previous studies using a soluble E-selectin-IgG chimera<sup>28</sup>.

Mutations that impair E-selectin adhesion to neutrophils fall into two categories based on the severity of the defect: either complete or partial ablation of binding. Both groups of mutations are clustered on the top face of the lectin domain in the vicinity of the bound  $\text{Ca}^{2+}$  (Fig. 4a-c), which is consistent with the finding that selectin-mediated neutrophil adhesion is  $\text{Ca}^{2+}$ -dependent<sup>30</sup>. Mutations at Asn 82, Tyr 94, Arg 97 and Lys 113 comprise the group that completely abolish neutrophil binding (Table 2). Among these, Asn 82 is a calcium ligand and is equivalent to Asn 187 in MBP (Fig. 2c). The structure of MBP complexed with an oligosaccharide<sup>26</sup> reveals that Asn 187 not only coordinates calcium but also donates a hydrogen bond through its amide moiety to a ligand hydroxyl group. Given these results, and the similarity of fucose and mannose, the analogy between the binding by the two proteins was tested by making the N82D (Asn→Asp at residue 82) substitution in E-selectin. As predicted, calcium still bound, as evidenced by resistance to protease digestion (ref. 26, and K.-S.H., unpublished results), but neutrophil adhesion was lost (Table 2). Thus, Asn 82 may have a role in binding  $\text{sLe}^x$  that is similar to the MBP Asn 187 in binding mannose.

Mutational analysis of MBP indicated that six of seven mutations that eliminated sugar binding were directly involved in  $\text{Ca}^{2+}$  binding<sup>31</sup>. In contrast, other residues in E-selectin identified as critical for neutrophil adhesion are not directly involved in ligating  $\text{Ca}^{2+}$ . Tyr 94 and Arg 97, which lie in the loop connecting  $\beta 4$  and  $\beta 5$ , as well as Lys 113 in the lectin domain C-terminal  $\beta$ -strand ( $\beta 6$ ), are essential for neutrophil adhesion. Arg 97 was very sensitive to substitution, as evidenced by the diminished cell adhesion with even the conservative substitution of a lysine residue. The most striking mutation, Y94F, results in the loss of just a single hydroxyl group but eliminates cell binding. The ring of Tyr 94 is largely buried under Arg 97 and blocked on the side by the side chains of Tyr 48 and Asp 100. The hydroxyl group of Tyr 94 makes no direct hydrogen bonds to any other part of the protein but does form two hydrogen bonds to water molecules (2.94 Å to both). Thus, the loss of the hydroxyl group in the Y94F mutant would not be expected to disrupt local protein structure and would therefore suggest that it does form one or more hydrogen bonds to the carbohydrate ligand. Arg 97 also forms a strong hydrogen bond to the side chain of Asp 100. Lys 113 is an interesting residue in that the  $-\text{NH}_3^+$  group forms an ion pair (2.98 Å) with Glu 92 while the alkyl portion of the side chain forms a hydrophobic contact with the buried Trp 50. Thus, Lys 113 would appear to have a dual role in the E-selectin structure and may interact directly with  $\text{sLe}^x$  as well.

Another group of mutations was identified with a less severe impact on adhesion *in vitro*. The mutants Y48F, N83A, E92A and K111A formed rosettes with neutrophils during the binding

TABLE 2 Neutrophil adhesion of E-selectin mutants

|           |   |           |   |                           |    |
|-----------|---|-----------|---|---------------------------|----|
| Tyr94Phe  | — | Gln30Asn  | + | Glu8Ala                   | +  |
| Arg97Lys  | — | Glu33Lys  | + | Arg84Ala/Lys86Ala         | ++ |
| Arg97Gln  | — | Glu36Lys  | + | Glu8Ala/Arg84Ala/Lys86Ala | ++ |
| Arg97Ala  | — | Tyr44His  | + | Arg84Ala                  | ++ |
| Lys113Ala | — | Tyr44Phe  | + | Arg84Lys                  | ++ |
| Asn82Asp  | — | Asn57Ala  | + | Lys86Ala                  | +  |
|           |   | Asn58Ala  | + |                           |    |
| Tyr48Phe  | ± | Gln85Ala  | + | Glu107Ala/Lys111Ala       | +  |
| Asn83Ala  | ± | Asp87Ala  | + | Arg97Ala/Asp100Ser        | —  |
| Glu92Ala  | ± | Asp89Ala  | + |                           |    |
| Lys111Ala | ± | Glu107Ala | + |                           |    |

Relative neutrophil adhesion of E-selectin mutants: +, wild-type level binding; —, no adhesion; ±, weak adhesion lost on washing; ++, greater than wild-type level of adhesion. Expression of E-selectin mutants on the surface of COS cells was achieved by fusing cDNA sequences encoding residues 1–532 of human E-selectin with sequences of CD16 sufficient to anchor the protein to the cell membrane by a glycosyl inositol phosphate linkage<sup>47</sup>. These sequences were cloned into the vector pEF-BOS, which places the E-selectin sequences under the control of the EF-1 $\alpha$  promoter<sup>48</sup>. Point mutations were introduced into the lectin domain according to ref. 49 using the Muta-Gene Phagemid *in vitro* Mutagenesis Kit (BioRad). To confirm the presence of the mutation and eliminate the possibility of random mutations, the DNA sequence of the lectin and EGF domains of each mutant was determined using an ABI model 373A sequencer with fluorescent chain terminators. COS cells were transfected with the E-selectin constructs by the DEAE-dextran method and 44–72 h later used for structure/function determination. Surface expression of the E-selectin mutant was verified using monoclonal antibody (mAb) 9A1, which recognizes an epitope within cr1/cr2. FACS analysis with mAb 9A1 demonstrated that although transfection efficiency varied, an equivalent amount of protein was expressed on the cell surface for each mutant. The structural integrity of each mutant was assessed by immunofluorescence with mAbs 3B7, 8E4, 7H5, 1D6, 1E5 and 14G2, which recognize determinants within the lectin and EGF domains<sup>28</sup>. 3B7, 8E4, and 7H5 block E-selectin-dependent adhesion of neutrophils. Human polymorphonuclear cells were isolated from fresh whole blood, loaded with 6-carboxyfluorescein diacetate, and incubated with transfected COS cells for 20 min at 25 °C<sup>28</sup>. Non-adherent cells were removed by aspiration, and the ability of each mutant to mediate neutrophil adhesion was determined by visual inspection and photographed. COS cells transfected with vector only sequences served as the negative control in each experiment. To quantify adhesion, cells were briefly incubated in trypsin, and the amount of fluorescence associated with adherent neutrophils was determined with a CytoFluor model 2300 plate reader (Millipore).

phase of the adhesion assay. These bound neutrophils, however, were readily removed by washing. Interestingly, these residues are located in the vicinity of the null mutants, with their side chains extending in the direction of the top face of the lectin domain (Fig. 4a-c). Furthermore, two of these interact directly with residues characterized as null mutants: Tyr 48 forms an aromatic-aromatic interaction<sup>32</sup> with Tyr 94, and Glu 92 forms an ion pair with Lys 113. Glu 92 also forms a hydrogen bond with the amide group of Asn 105, which is a calcium ligand. The only interaction in which Lys 111 seems to participate is to form an ion pair with Glu 107. Thus, whether Lys 111 makes a weak but specific contact with the ligand, or whether the alanine substitution has an effect on local structure and/or calcium binding is unknown.

The lec/EGF domains of E-selectin are not generally sensitive to mutagenesis because several substitutions had no effect on neutrophil adhesion or recognition by a panel of monoclonal antibodies (Table 2; and ref. 28). This suggests that the residues identified as critical for function are directly involved in ligand recognition. Interestingly, a few amino-acid substitutions were identified that increased neutrophil binding (Table 2). The mechanism for this enhanced adhesion is not evident from the

structure of E-selectin and will require a better understanding of the native ligand. In support of our adhesion assay results, the amino-acid residues identified as critical for neutrophil adhesion (Tyr 48, Asn 82, Asn 83, Glu 92, Tyr 94, Lys 111 and Lys 113) are conserved among all members of the selectin family<sup>3,5,33,37</sup>. In addition, mutagenesis studies of P-selectin indicate that amino-acid residues at positions 48, 94, 111 and 113 are also critical for binding the myeloid cell line HL-60 (refs 38, 39).

### sLe<sup>x</sup> binding to E-selectin

The available structural and mutational data on E-selectin were evaluated in the light of the known solution NMR structure of sLe<sup>x</sup> (refs 40–43). The residues identified (Fig. 4a–c) define a limited and feasible site for sLe<sup>x</sup> binding. MBP complexed with

mannose reveals that the sugar binds directly to the calcium ion, with additional hydrogen bonds to some of the protein calcium ligands<sup>26</sup>. In addition to the hydrogen bond formed with Asn 187, mannose also donates two hydrogen bonds to Glu 185 and Glu193 and accepts another from Asn 205. In E-selectin, Asn 82, Glu 80 and Asn 105 are conserved both in identity (Asn 187, Glu 185 and Asn 205, respectively) and function in binding calcium relative to MBP, but Glu 88 (equivalent to Glu 193 in MBP), although identical, does not bind calcium and seems unable to form the same hydrogen bond to sLe<sup>x</sup>. Nevertheless, sLe<sup>x</sup> docks to the E-selectin surface when fucose is in a configuration relative to the calcium ion similar to that of mannose bound to MBP. The NMR structure of sLe<sup>x</sup> can be inserted into this region of E-selectin without severe van der

FIG. 1 E-selectin structure. *a*, Stereoview of a segment of the original multiple isomorphous replacement (MIR) electron density map with the final refined protein and solvent models included. The map is calculated from phases derived from gadolinium and platinum derivatives (R.L.C. and B.J.G., manuscript in preparation) for data to 2.5 Å resolution. The map is contoured at 1.0 $\sigma$ . Most of the protein structure depicted is from the first helix with Tyr 18 appearing prominently in the middle. Strong density is exhibited throughout for the main chain and most side chains and density is apparent for some solvent molecules. The strong density connected to the main chain at Gln 20 was determined to be a Ca<sup>2+</sup> ion and several coordinated water molecules. Based on this MIR map, *C $\alpha$*  positions were established using FRODO<sup>50</sup> and the *C $\alpha$*  model was converted into a full-atom structure using the molecular modelling package SYBYL<sup>51</sup>. This full-atom model was subsequently rebuilt in FRODO and used as the starting point for refinement. *b* and *c*, Ribbon drawings of lec/EGF E-selectin in two orientations. The lec domain is depicted by the blue ribbon and the EGF domain by the magenta ribbon. The three calcium ion positions are shown as white spheres and the ten cysteines that form five disulphide bridges are shown as yellow sticks. *b*, The molecule in this orientation will be referred to as the 'front' view. The main orienting feature is that this view looks down the helix axis of the second  $\alpha$ -helix. This view also shows the weak binding sites of two of the calcium ions—one at the C terminus of the first  $\alpha$ -helix ( $\alpha$ 1) and the other bound to two glutamic acid residues (not shown) in the second  $\alpha$ -helix ( $\alpha$ 2). Labels for strands  $\beta$ 2 and  $\beta$ 3 are omitted in this view as the strands are in the back and not clearly visible. *c*, The same picture rotated roughly 90°, referred to as the 'side' view, reveals the distinct relationship between the lec and EGF domains.

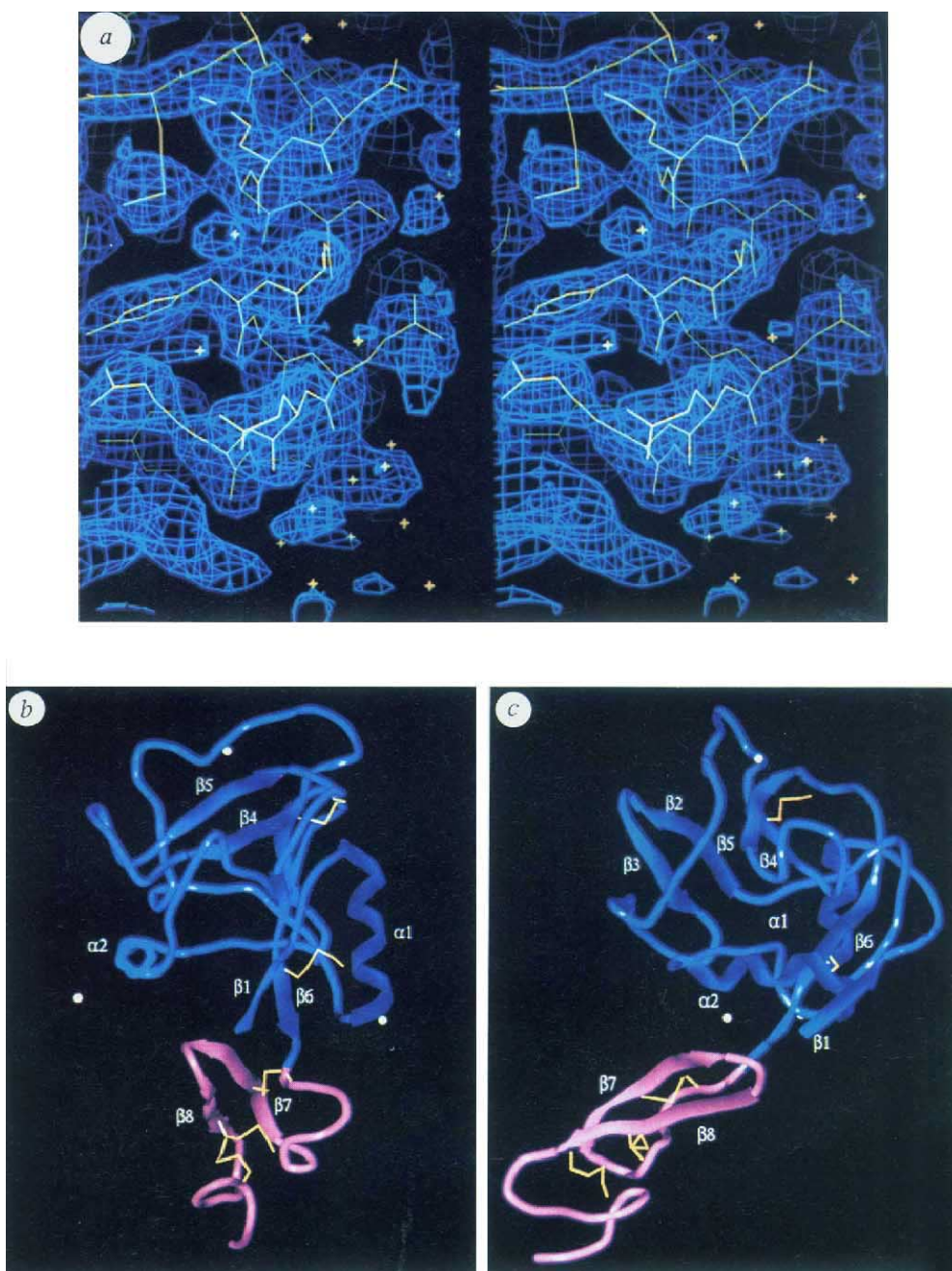
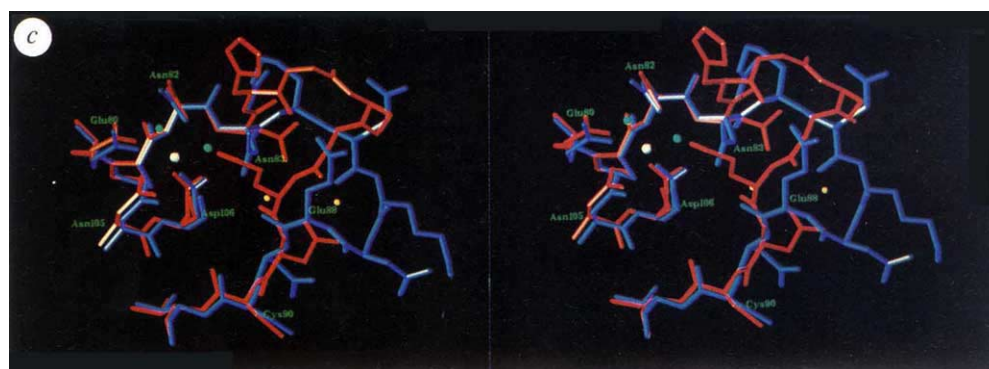
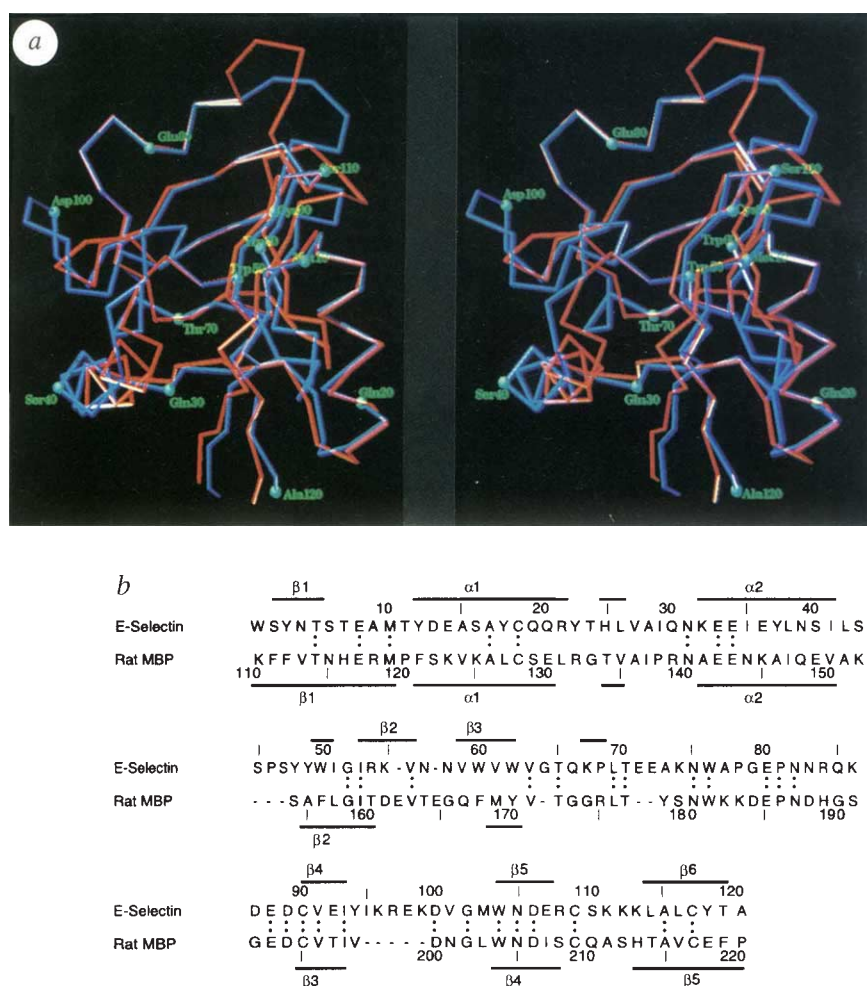




FIG. 2 Comparison of E-selectin and MBP lectin domains. *a*, Stereoview of the overlay of the  $C\alpha$  structures of E-selectin (blue) and MBP<sup>26</sup> (red) lectin domains. The overlay was achieved by means of a best molecular fit using 109 topologically equivalent  $C\alpha$  positions in the two proteins. Every tenth residue in the E-selectin lectin domain, starting with Met 10, is denoted with a cyan sphere and labelled. *b*, Topological alignment of the lectin domains. Alignment was accomplished by comparing the overlaid structures (a best molecular fit based on matching just the identical residues) and identifying the looped-out residues. This allowed the identification of all 109 common pairs. The colons between the sequence lines denote identical residues. There are 33 identities in 109 pairs (30.3%). Sequences for MBP outside the termini of the E-selectin lectin domain are not shown. Assignment of secondary structure elements was done using program DSSP with the Kabsch and Sander routines<sup>52</sup>. Secondary structure elements are indicated by bars and are labelled. *c*, Stereoview of the 'conserved' calcium-binding site in E-selectin and in MBP<sup>26</sup>. The E-selectin model, shown in blue with a white sphere representing calcium and blue spheres representing the two water molecules bound to calcium, includes residues Glu 80–Val 91 and Asn 105–Asp 106. The corresponding residues from MBP (Glu 185–Val 196 and Asn 205–Asp 206) are shown in red, with the three calcium ions as yellow spheres. The 'conserved' calcium ion in the MBP structure is hidden behind the white sphere of the calcium ion from E-selectin. The three-dimensional alignment depicted here was calculated using the 109 topologically equivalent  $C\alpha$  positions in the lectin domains of E-selectin and MBP. Residue labels correspond to the E-selectin numbering.



Waals clashes. One possible scheme (Fig. 4d) positions two fucose hydroxyl groups so that they replace the two water molecules that bind to the calcium ion and the remainder of the sLe<sup>x</sup> structure so that it extends towards Lys 113, but this model is speculative and sLe<sup>x</sup> could well bind at a different site or in a different way. The carbohydrate ligand may even be a dimeric form of sLe<sup>x</sup> (ref. 44), although the increase in binding for dimeric sLe<sup>x</sup> is only small. Nonetheless, Fig. 4d shows that the residues important for sLe<sup>x</sup> binding, as defined by mutagenesis, as well as the calcium ion, can all be encompassed by the sLe<sup>x</sup> structure.

## Discussion

The structure of the lec/EGF domains of human E-selectin reported here will allow further analysis of selectin-dependent

cell adhesion. Although both the lec and EGF domains are required to mediate neutrophil adhesion, the structure reveals two distinct modules with limited association. In addition, although the overall fold of the E-selectin lectin domain is similar to MBP<sup>18,26</sup>, there are differences, particularly in the inserted loops and in the amino-acid residues involved in coordinating Ca<sup>2+</sup>, which have important implications for function. Mapping of amino-acid substitutions that eliminate neutrophil adhesion onto the E-selectin structure identifies a finite region of the lectin domain in the vicinity of the bound Ca<sup>2+</sup> as the most likely carbohydrate contact surface. These features of the interaction between E-selectin and its ligand offer new opportunities for the design of compounds to regulate the accumulation of leukocytes at sites of inflammation. □





Received 1 November; accepted 30 December 1993.

1. Bevilacqua, M. P., Prober, J. S., Mendrick, D. L., Cotran, R. S. & Gimbrone, M. A. Jr *Proc. natn. Acad. Sci. U.S.A.* **84**, 9238–9242 (1987).
2. Siegelman, M. H., van de Rijn, M. & Weissman, I. L. *Science* **243**, 1165–1172 (1989).
3. Siegelman, M. H. & Weissman, I. L. *Proc. natn. Acad. Sci. U.S.A.* **86**, 5562–5566 (1989).
4. Lasky, L. A. et al. *Cell* **56**, 1045–1055 (1989).
5. Johnston, G. I., Cook, R. G. & McEver, R. P. *Cell* **56**, 1033–1044 (1989).
6. Butcher, E. C. *Cell* **67**, 1033–1036 (1991).
7. Lasky, L. A. *Science* **258**, 964–969 (1992).
8. Bevilacqua, M. P. *Rev. Immun.* **11**, 767–804 (1993).
9. Lasky, L. A. et al. *Cell* **69**, 927–938 (1992).
10. Moore, K. L. et al. *J. Cell Biol.* **118**, 445–456 (1992).
11. Norgard, K. E. et al. *J. biol. Chem.* **268**, 12764–12774 (1993).
12. Levinovitz, A., Muhlhoff, J., Isenmann, S. & Vestweber, D. *J. Cell Biol.* **121**, 449–459 (1993).
13. Foxall, C. et al. *J. Cell Biol.* **117**, 895–902 (1992).
14. Bowen, B. R., Fennie, C. & Lasky, L. A. *J. Cell Biol.* **110**, 147–153 (1990).
15. Walz, G., Aruffo, A., Kolanus, W., Bevilacqua, M. & Seed, B. *Science* **250**, 1132–1135 (1990).
16. Pigott, R., Needham, L. A., Edwards, R. M., Walker, C. & Power, C. J. *Immunology* **147**, 130–135 (1991).
17. Li, S. H. et al. *J. biol. Chem.* (in the press).
18. Weis, W. I., Kahn, R., Fourme, R., Drickamer, K. & Hendrickson, W. A. *Science* **254**, 1608–1615 (1991).
19. Hommel, U., Harvey, T. S., Driscoll, P. C. & Campbell, I. D. *J. molec. Biol.* **227**, 271–282 (1992).
20. Montellione, G. T. et al. *Biochemistry* **31**, 236–249 (1992).
21. Kohda, D., Shimada, I., Miyake, T., Fuwa, T. & Inagaki, F. *Biochemistry* **28**, 953–958 (1989).
22. Kline, T. P. et al. *Biochemistry* **29**, 7805–7813 (1990).
23. Harvey, T. S., Wilkinson, A. J., Tappin, M. J., Cooke, R. M. & Campbell, I. D. *Eur. J. Biochem.* **198**, 555–562 (1991).
24. Baron, M. et al. *Prot. Sci.* **1**, 81–90 (1992).
25. Selander-Sunnerhagen, M. et al. *J. biol. Chem.* **267**, 19642–19649 (1992).
26. Weis, W. I., Drickamer, K. & Hendrickson, W. A. *Nature* **360**, 127–134 (1992).
27. Bezouska, K., Crichtlow, G. V., Rose, J. M., Taylor, M. E. & Drickamer, K. *J. biol. Chem.* **266**, 11604–11609 (1991).
28. Erbe, D. V. et al. *J. Cell Biol.* **119**, 215–227 (1992).
29. Padmanabhan, K. et al. *J. molec. Biol.* **232**, 947–966 (1993).
30. Hession, C. et al. *Proc. natn. Acad. Sci. U.S.A.* **87**, 1673–1677 (1990).
31. Quesenberry, M. S. & Drickamer, K. *J. biol. Chem.* **267**, 10831–10841 (1992).
32. Hunter, C. A., Singh, J. & Thornton, J. M. *J. molec. Biol.* **218**, 837–846 (1991).
33. Bevilacqua, M. P., Stengelin, S., Gimbrone, M. A. Jr & Seed, B. *Science* **243**, 1160–1165 (1989).
34. Larigan, J. D., Tsang, T. C., Rumberger, J. M. & Burns, D. K. *DNA Cell Biol.* **11**, 149–162 (1992).
35. Becker-Andre, M., Van Huijsduijnen, R. H., Losberger, C., Whelan, J. & Delamarier, J. F. *Eur. J. Biochem.* **206**, 401–411 (1992).
36. Weiler, A., Isenmann, S. & Vestweber, D. *J. biol. Chem.* **267**, 15176–15183 (1992).
37. Strubel, N. A., Nguyen, M., Kansas, G. S., Tedder, T. F. & Bischoff, J. *Biochem. biophys. Res. Commun.* **192**, 338–344 (1993).
38. Erbe, D. V. et al. *J. Cell Biol.* **120**, 1227–1235 (1993).
39. Hollenbaugh, D., Bajorath, J., Stenkamp, R. & Aruffo, A. *Biochemistry* **32**, 2960–2966 (1993).
40. Wormald, M. R., Edge, C. J. & Dwek, R. A. *Biochem. biophys. Res. Commun.* **180**, 1214–1221 (1991).
41. Ball, G. E. et al. *J. Am. chem. Soc.* **114**, 5449–5451 (1992).
42. Lin, Y.-C. et al. *J. Am. chem. Soc.* **114**, 5452–5454 (1992).
43. Miller, K. E., Mukhopadhyay, C., Cagas, P. & Bush, C. A. *Biochemistry* **31**, 6703–6709 (1992).
44. DeFrees, S. A., Gaeta, F. C. A., Lin, Y.-C., Ichikawa, Y. & Wong, C.-H. *J. Am. chem. Soc.* **115**, 7549–7550 (1993).
45. Brünger, A. T., Kuriyan, J. & Karplus, M. *Science* **235**, 458–460 (1987).
46. Brünger, A. T. *Nature* **355**, 472–475 (1992).
47. Scallan, B. J., Kado-Fong, H., Nettleton, M. Y. & Kochan, J. P. *Biotechnology* **10**, 550–556 (1992).
48. Mizushima, S. & Nagata, S. *Nucleic Acid Res.* **18**, 5322 (1990).
49. Kunkel, T. A., Roberts, J. D. & Zakour, R. A. *Meth. Enzym.* **154**, 367–382 (1987).
50. Jones, T. A. *J. appl. Crystallogr.* **11**, 268–272 (1978).
51. SYBYL: Molecular Modeling Software, Version 5.9 (Tripos Associates, St Louis, MO, 1992).
52. Kabsch, W. & Sander, S. *Biopolymers* **22**, 2577–2637 (1983).

ACKNOWLEDGEMENTS. We gratefully acknowledge W. McComas and R. Motyka for synthesis of oligonucleotides, J. D. Larigan and J. Levine for DNA sequencing, D. Gaizband for neutrophils, S. Nagata for the expression plasmid pEF-BOS, J. Kochan for the E-selectin/CD16 fusion construct, E. Rios-Fredrickson for assistance with manuscript preparation, and the Cell Adhesion Project Team for support and discussion.

## LETTERS TO NATURE

# A possible local counterpart to the excess population of faint blue galaxies

Stacy S. McGaugh

Institute of Astronomy, University of Cambridge, Madingley Road, Cambridge CB3 0HA, UK

OBSERVATIONS have revealed a population of faint blue galaxies<sup>1–5</sup> at intermediate redshifts ( $z \approx 0.4$ ). These galaxies are present in significant excess relative to what is expected based on observations of local galaxies, which has led some to propose that the discrepancy must be due either to non-standard cosmologies<sup>4,6</sup> or to the effects of very pronounced galaxy evolution<sup>4,7–10</sup>. The surveys that define the population of local galaxies are strongly biased against objects of low surface brightness<sup>11–14</sup>. Recent work<sup>15,16</sup> has shown that nearby low-surface-brightness galaxies have properties very similar to those of the excess faint blue galaxies, and has suggested that they may be as common as normal spiral galaxies<sup>14,17</sup>. Here I show that the deep surveys that have identified the faint blue galaxies can easily detect low-surface-brightness galaxies at intermediate redshifts, even though they are not readily apparent in local surveys. Thus, the faint blue galaxies may indeed correspond to low-surface-brightness galaxies.

Proposed solutions to the discrepancy between the number counts of galaxies in the B and I bands<sup>1–5</sup>, predicted using standard cosmological models ( $0 < q \leq 0.5$ , with little or no galaxy evolution for  $z < 0.5$ ), have involved altering either cosmology or the galaxy population. The former changes the geometry of the Universe<sup>4,6</sup> in order to reconcile the high counts at intermediate redshifts with the low counts at low redshifts, whereas the latter supposes strong evolution in the numbers of galaxies<sup>4,7</sup>,

selective evolution in their luminosity<sup>8,9</sup>, or an entirely new population<sup>10</sup>. The most common alteration of cosmology invokes a finite cosmological constant<sup>6</sup>, but this may already be in conflict with limits imposed by the statistics of gravitational lenses<sup>18</sup>. Evolution in the numbers of galaxies provides a good description of the observations<sup>4,7,19</sup>, but the high rate of mergers necessary at  $z \approx 0.1$  is difficult to reconcile with the low current merger rate and limits on the mass accreted by spiral disks<sup>20</sup>. It also goes against the observation that the faint blue galaxies are in relatively isolated environments<sup>10,21</sup>, rather than the dense regions where merging is more likely to occur. Luminosity evolution must be selective<sup>8</sup> in order to match the apparent evolution in number density. In this hypothesis, most galaxy types evolve mildly, but the dwarf galaxies which are common and faint at the present epoch must have evolved strongly<sup>8,9</sup> so that there was a high density of bright objects at the appropriate redshift. The amount of evolution required is extreme, but has not been observed<sup>22</sup>. None of these explanations appears satisfactory, suggesting that the excess faint blue galaxies at moderate redshift may have a local counterpart that has remained unidentified until now.

A population of galaxies with physical properties similar to those of the faint blue galaxy population does exist: low-surface-brightness disk galaxies. These galaxies have central surface brightnesses,  $\mu_0$ , well below what is considered<sup>23</sup> 'normal' for spirals ( $\mu_0^B \geq 23$  mag arcsec<sup>-2</sup> rather than  $\mu_0^B \approx 21.6$  mag arcsec<sup>-2</sup>; B indicates B-band). They have colours and luminosities<sup>16,24</sup> comparable to those of the excess population (Fig. 1), so invoking extreme forms of evolution is not necessary. Indeed, the star formation history inferred for low-surface-brightness galaxies<sup>16</sup> suggests that they should not change much in colour and luminosity for  $z < 0.4$ . They remain blue because they form stars at a low, approximately constant rate, and have not had time to build up a substantial red giant branch since their relatively recent epoch of formation (very roughly,  $z \sim 1$ ).

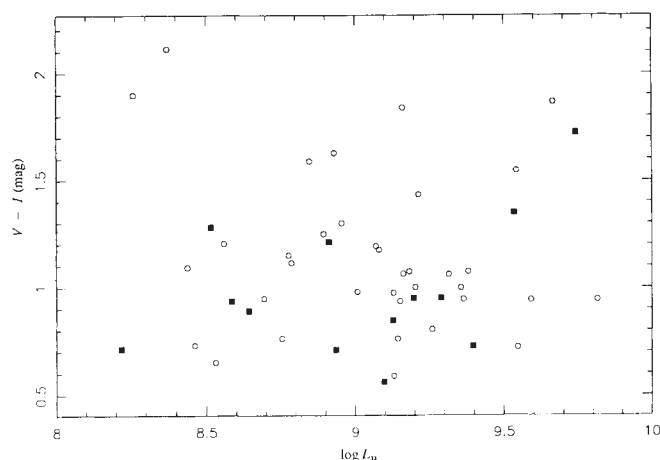


FIG. 1 The V-I colour plotted against the log of the blue luminosity ( $L_B$ ) for a sample of faint<sup>8</sup> (open circles) and low-surface-brightness<sup>16</sup> (filled squares) galaxies. The faint galaxy sample shows the expected number of red galaxies, but an excess in the number of galaxies with blue colours ( $V-I \approx 1$ ). The faint-galaxy data have been shifted to the rest frame following ref. 8 and assuming that  $H_0 = 100 \text{ km s}^{-1} \text{ Mpc}^{-1}$ , where  $H_0$  is Hubble's constant relating redshift to distance. The excess faint blue galaxies occupy the same region of this diagram as do low-surface-brightness galaxies.

In addition, low-surface-brightness disks are observed to be weakly clustered<sup>17,25</sup>, with a correlation amplitude lower than normal galaxies by about the same factor as the faint galaxy population<sup>17,21</sup>. Other observed properties of the excess faint population, such as star formation<sup>7,26</sup> and morphology<sup>8</sup>, though difficult to quantify, appear to be consistent with those of low-surface-brightness galaxies<sup>15</sup>. This correspondence of physical properties makes it natural to identify the two populations with one another.

For this identification to be correct, the deep surveys that reveal the excess population must be more sensitive to low-surface-brightness galaxies than local surveys. At first this may seem counterintuitive, as cosmological  $(1+z)^4$  dimming (due to the noneuclidean geometry of the universe) will decrease initially low surface brightness even further. This applies equally to all galaxies, however.

Surveys are usually said to be complete to some limiting flux, but strictly speaking this is only correct in the case of point sources. As galaxies are resolved, it is also necessary to specify the isophotal threshold above which the flux is measured. The portion of a galaxy that is detected depends on the luminosity profile ( $\mu(r)$ ) of that galaxy. The types of galaxies of interest here can be described as exponential disks with profiles of the form

$$\mu(r) = \mu_0 + 1.086 \frac{r}{l}$$

The two parameters  $\mu_0$  and  $l$  (central surface brightness and exponential scale length) characterize the surface brightness distribution of the disk of a galaxy and allow us to determine the

ratio of observed flux ( $F_{\text{obs}}$ ) to total flux ( $F_{\text{tot}}$ ) that is actually measured by a survey. This ratio is

$$F_{\text{obs}}/F_{\text{tot}} = 1 - (1+n)e^{-n}$$

where  $n$  is the number of scale lengths  $l$  observed above the isophotal limit. This convenient analytical approximation, which will slightly underestimate the observed fraction of light from spirals with bulges, is nevertheless an important improvement over treating galaxies as point sources.

The isophotal aperture, in scale lengths  $n$ , of a survey conducted to a limiting surface brightness  $\mu_1$  is given by

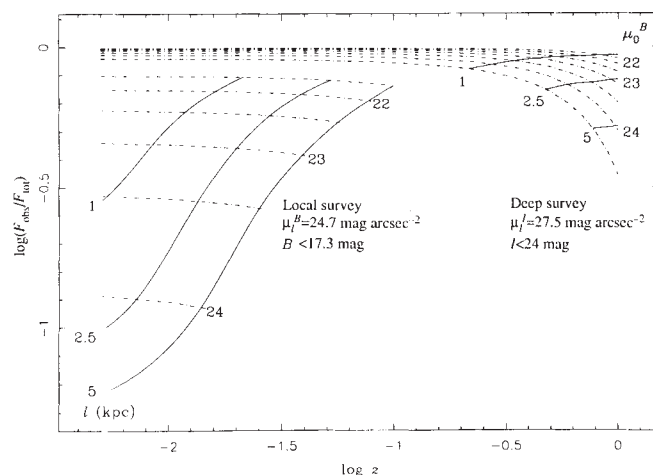
$$n = \frac{\mu_1 - \mu_0 - 10 \log(1+z) - k(z)}{1.086}$$

The terms involving  $z$  account for  $(1+z)^4$  surface brightness dimming and the redshifting of the spectral energy distribution (the  $k$ -correction<sup>27</sup>). The  $k$ -corrections depend on galaxy type as well as redshift, but are negligibly small for the blue galaxies of interest here.

The great sensitivity of the deep surveys more than overcomes  $(1+z)^4$  dimming and low-surface-brightness galaxies can be seen to intermediate redshifts because essentially all the flux is detected (Fig. 2). For the typical observed redshift ( $z \sim 0.4$ ),  $F_{\text{obs}}/F_{\text{tot}} > 0.8$  for  $\mu_0^B < 24 \text{ mag arcsec}^{-2}$ , and even modest-sized galaxies exceed the flux limit. Thus, detecting low-surface-brightness galaxies at intermediate redshift is not a problem.

It is rather more of a problem locally. Normal spirals<sup>23</sup> can be detected to distances  $\sim 3$  times as great as low-surface-brightness disks of the same size, and hence sample a volume  $\sim 27$  times larger. This causes the numbers of low-surface-brightness galax-

FIG. 2 Log-log plot of the fraction of the total flux ( $F_{\text{tot}}$ ) observed ( $F_{\text{obs}}$ ) within isophotal apertures as a function of redshift ( $z$ ) for galaxies with exponential profiles. The specific case of the selection characteristics of the deep survey of ref. 4 is shown in the upper right; that for a large-area local survey<sup>27,35</sup>, which is commonly taken to define the luminosity function<sup>28</sup>, occupies the left part of the diagram. The isophotal ( $\mu_1$ ) and flux limits (in the appropriate band) are indicated in the figure. Dash-dotted lines of constant rest frame central surface brightness (in half-magnitude steps labelled by  $\mu_0^B$ ) decline because of cosmological  $(1+z)^4$  dimming. Solid lines delimit the distance to which galaxies of different sizes can be seen for the flux limits of the two surveys, including the effect of flux lost outside the aperture ( $F_{\text{obs}}/F_{\text{tot}} < 1$ ). The solid lines are labelled by the size characteristic  $l$ , the exponential scale length. The two surveys have very different selection characteristics. The deep survey detects nearly all the flux of low-surface-brightness galaxies even at large redshifts, whereas the local survey detects only a fraction of their total flux. This causes low-surface-brightness galaxies to be missed locally; observing a population comparable to that detected by deep surveys would require a local survey with an isophotal limit more sensitive by  $\sim 2.5 \text{ mag}$ .





ies to be seriously underestimated. Even when such galaxies are specifically selected, isophotal measures systematically underestimate the flux of low-surface-brightness systems. This effect is severe; for  $\mu_0^B = 23 \text{ mag arcsec}^{-2}$ ,  $F_{\text{obs}}/F_{\text{tot}} \approx 0.45$ , which drops  $F_{\text{obs}}/F_{\text{tot}} \approx 0.12$  for  $\mu_0^B = 24 \text{ mag arcsec}^{-2}$ . This causes systematic errors in the derived luminosity function and number counts<sup>28,29</sup>—the luminosity will be too flat, and the number counts too steep. The steep slope of the number counts at relatively bright magnitudes<sup>29</sup> is incompatible with any reasonable amount of galaxy evolution for the short time since  $z \sim 0.1$ , but is expected from the use of isophotal magnitudes when the distribution of galaxy surface brightnesses is not a delta function.

The extended nature of galaxies means that it is necessary to consider the bivariate distribution of luminosity and surface brightness rather than the luminosity function (the space density of galaxies of luminosity  $L$ ) alone. The bivariate distribution is not well quantified among isolated galaxies<sup>30</sup>, but in clusters the surface brightness distribution is observed to be flat<sup>31,32</sup> rather than sharply peaked<sup>23</sup>. To illustrate the effect that this will have on the luminosity function derived by various surveys, the expression for  $F_{\text{obs}}/F_{\text{tot}}$  can be combined with assumed forms of the bivariate distribution (Fig. 3). The luminosity function ( $\phi$ ) is usually taken to be of the form<sup>33</sup>

$$\phi(L) = \phi^* \left(\frac{L}{L^*}\right)^\alpha e^{-L/L^*}$$

where  $\phi^*$  describes the density of galaxies,  $L^*$  is a characteristic luminosity brighter than which galaxies are rare ( $L_B^* \sim 10^{10} h^{-2} L_\odot$ ), and  $\alpha$  is the asymptotic slope of the faint end. These parameters can be seriously mistaken by standard pro-

cedures which implicitly assume that  $F_{\text{obs}}/F_{\text{tot}}$  is a constant. In particular, local surveys are prone to underestimating the numbers of objects with  $L < \frac{1}{2} L^*$ , which is where the discrepancy with faint counts occurs.

Specifically, it is possible to find many bivariate distributions that reproduce the locally determined luminosity function and predict an apparent excess in the numbers of objects seen in deep surveys. If, as seems likely, the true bivariate distribution is intermediate between the extreme cases illustrated in Fig. 3, so that there is an average trend of luminosity with surface brightness with broad scatter, then both the density of galaxies  $\phi^*$  and the slope  $|\alpha|$  (which describes the frequency of faint galaxies) will be underestimated locally. This may be sufficient to explain the entire discrepancy in the faint galaxy number counts.

Identification of the faint galaxy excess with low-surface-brightness galaxies is actually a conservative hypothesis, as no drastic alterations to cosmology or radical forms of galaxy evolution are required. Large numbers of low-surface-brightness galaxies are already indicated locally by their high surface number density<sup>14</sup> and the simple model which matches their spatial distribution<sup>17</sup>. Because isophotal measures are inadequate when a distribution of surface brightness is present, it is necessary to measure the distribution of both surface brightness and luminosity, rather than leaping directly to the luminosity function.

Although galaxy evolution inevitably occurs, it will be impossible to determine from number counts, much less derive cosmological parameters, until the bivariate distribution is quantified. Indeed, correcting for redshift effects requires detailed knowledge of the spectral energy distribution (SED), so interpreting data on galaxies at significant redshift requires knowledge of the

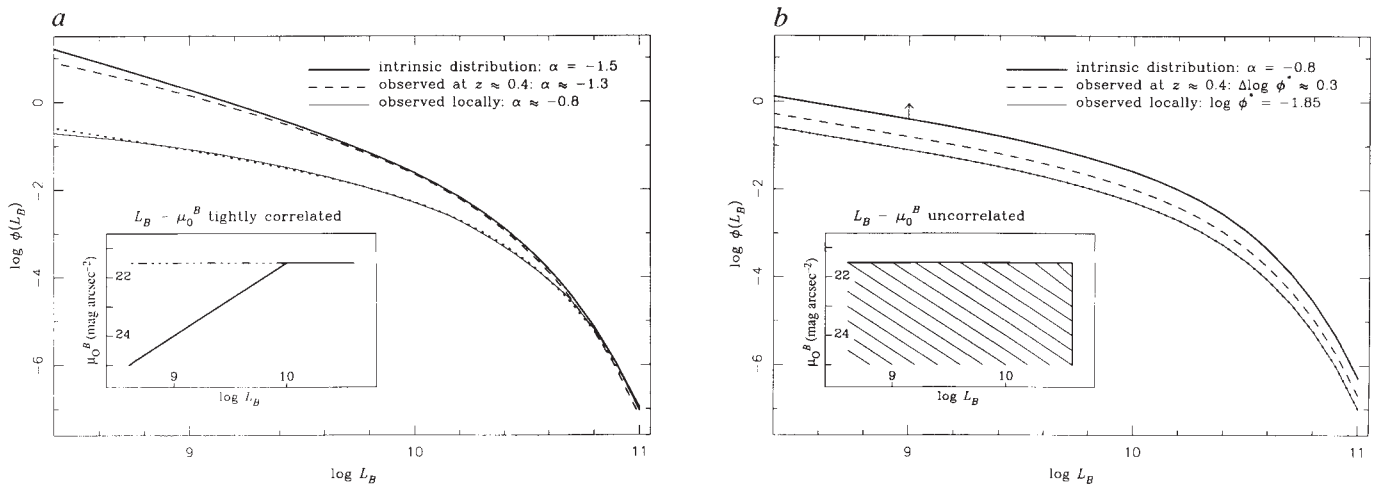


FIG. 3 Log-log plot of the luminosity function  $\phi(L_B)$  against the blue luminosity  $L_B$ , derived from two forms of the bivariate galaxy distribution. **a**, A bivariate distribution in which luminosity ( $L_B$ ) and surface brightness ( $\mu_0^B$ ) are tightly correlated. **b**, No correlation between luminosity and surface brightness (see insets). These opposite forms of the bivariate distribution (the true form of which is not well known) illustrate the hazards posed by isophotal measures and surface brightness selection effects. In **a**, surface brightness falls off linearly with luminosity below  $L^*$  (solid line in inset). This is not a realistic bivariate distribution, but neither is one in which the surface brightness distribution is a delta function (the dashed/triple-dotted line in the inset, implicit in most determinations of the luminosity function which assume that  $F_{\text{obs}}/F_{\text{tot}}$  is a constant). For an intrinsic luminosity function with a slope  $\alpha = -1.5$  (the heavy solid line in the main figure), the dashed line will be observed by the deep survey of ref. 4 while the thin solid line will be observed by the local survey of ref. 28. These have been shifted in normalization to agree with the observations at the bright end: the dotted line is the luminosity function actually derived in ref. 28 for disk-type galaxies. The thin solid line provides as good a description of the observations as the dotted line, despite representing a very different intrinsic distribution.

This occurs because the luminosities of low-surface-brightness galaxies are systematically underestimated, leading to an underestimate of the slope  $|\alpha|$ , and a translation in  $L^*$  and  $\phi^*$ . The effect is more severe in the case of the local survey, causing an apparent excess in the numbers of galaxies observed by the deep survey (the difference between the dashed and thin solid line) even though both are drawn from the same intrinsic distribution. In **b**, galaxies fill the luminosity-surface brightness plane below a maximum<sup>12</sup> surface brightness without correlation (inset). This is a realistic bivariate distribution<sup>30-32</sup>, though not well quantified in the field. In this case, the observed  $\phi^*$  depends on the maximum surface brightness of the distribution and the limiting isophote of a survey. More sensitive surveys will probe further into the surface brightness distribution, again causing an apparent excess in the numbers of galaxies observed in deep relative to local surveys. There is no guarantee that there is a lower limit to the surface brightness distribution, hence the arrow on the intrinsic distribution. Detecting progressively fainter, higher-redshift galaxies necessarily requires achieving progressively greater isophotal sensitivity, so a bivariate of this sort predicts that the apparent  $\phi^*$  should increase with  $z$ , as observed<sup>19</sup>.

trivariate distribution function  $\Psi(L, \mu_0, \text{SED})$ . If low-surface-brightness galaxies maintain their blue colours into the ultra-violet, they may actually be the most prominent objects at  $z \approx 0.4$ .

Galaxy distributions of the sort suggested here can alleviate many outstanding problems in extragalactic astronomy which, like the faint blue galaxies, require more objects than provided by the apparent local luminosity function. Examples include the large number of  $Ly-\alpha$  absorption systems along the line of sight to quasars<sup>34</sup>, the baryonic mass missing relative to the otherwise successful predictions of primordial nucleosynthesis theory<sup>35</sup>, and the high measured value of the extragalactic background radiation<sup>36</sup>. Theories of cosmic structure formation should be used to predict the bivariate distribution, rather than the luminosity function alone. This may remove the need for *ad hoc* mechanisms to flatten the predicted luminosity function to match local observations. More fundamentally, we need to ask what is meant by frequently used qualitative terms like 'normal' galaxies. By this, do we mean a type of galaxy that is common (like faint main sequence stars), or one that merely happens to be optically prominent (like red giants)?  $\square$

Received 5 August; accepted 21 December 1993.

1. Tyson, A. J. *Astr. J.* **96**, 1–23 (1988).
2. Broadhurst, T. J., Ellis, R. S. & Shanks, T. *Mon. Not. R. astr. Soc.* **235**, 827–856 (1988).
3. Colles, M. M., Ellis, R. S., Taylor, K. & Hook, R. H. *Mon. Not. R. astr. Soc.* **244**, 408–423 (1990).
4. Lilly, S. J., Cowie, L. L. & Gardner, J. P. *Astrophys. J.* **369**, 79–105 (1991).
5. Metcalfe, N., Shanks, T., Fong, R. & Jones, L. R. *Mon. Not. R. astr. Soc.* **249**, 498–522 (1991).
6. Yoshii, Y. *Astrophys. J.* **403**, 552–566 (1993).
7. Broadhurst, T. J., Ellis, R. S. & Glazebrook, K. *Nature* **355**, 55–58 (1992).
8. Lilly, S. J. *Astrophys. J.* **411**, 501–512 (1993).
9. Babul, A. & Rees, M. J. *Mon. Not. R. astr. Soc.* **255**, 346–354 (1992).
10. Efsthathiou, G., Bernstein, G., Katz, N., Tyson, J. A. & Guhathakurta, P. *Astrophys. J.* **380**, L47–L50 (1991).
11. Disney, M. J. *Nature* **263**, 573–575 (1976).
12. Allen, R. J. & Shu, F. H. *Astrophys. J.* **227**, 67–79 (1979).
13. Bothun, G. D., Impey, C. D., Malin, D. F. & Mould, J. R. *Astr. J.* **94**, 23–29 (1987).
14. Schombert, J. M., Bothun, G. D., Schneider, S. E. & McGaugh, S. S. *Astr. J.* **103**, 1107–1133 (1992).
15. McGaugh, S. S. thesis, Univ. of Michigan (1992).
16. McGaugh, S. S. & Bothun, G. D. *Astr. J.* (in the press).
17. Mo, H. J., McGaugh, S. S. & Bothun, G. D. *Mon. Not. R. astr. Soc.* (in the press).
18. Moaz, D. & Rix, H.-W. *Astrophys. J.* **416**, 425–443 (1993).
19. Eales, S. *Astrophys. J.* **404**, 51–62 (1993).
20. Toth, G. & Ostriker, J. P. *Astrophys. J.* **389**, 5–11 (1992).
21. Roche, N., Shanks, T., Metcalfe, N. & Fong, R. *Mon. Not. R. astr. Soc.* **263**, 360–368 (1993).
22. Boroson, T. A., Salzer, J. J. & Trotter, A. *Astrophys. J.* **412**, 524–534 (1993).
23. Freeman, K. C. *Astrophys. J.* **160**, 811–820 (1970).
24. Schombert, J. M., Bothun, G. D., Impey, C. D. & Mundy, L. G. *Astr. J.* **100**, 1523–1531 (1990).
25. Bothun, G. D., Schombert, J. M., Impey, C. D., Sprayberry, D. & McGaugh, S. S. *Astr. J.* **106**, 530–547 (1993).
26. Koo, D. C. & Kron, R. G. *A. Rev. Astr. Astrophys.* **30**, 613–652 (1992).
27. Weinberg, S. *Gravitation and Cosmology* (Wiley, New York, 1972).
28. Loveday, J., Peterson, B. A., Efsthathiou, G. & Maddox, S. J. *Astrophys. J.* **390**, 338–344 (1992).
29. Maddox, S. J., Sutherland, W. J., Efsthathiou, G., Loveday, J. & Peterson, B. A. *Mon. Not. R. astr. Soc.* **247**, 1P–5P (1990).
30. Sodré, L. & Lahav, O. *Mon. Not. R. astr. Soc.* **260**, 285–298 (1993).
31. Irwin, M. J., Davies, J. I., Disney, M. J. & Phillips, S. *Mon. Not. R. astr. Soc.* **245**, 289–311 (1990).
32. Bothun, G. D., Impey, C. D. & Malin, D. F. *Astrophys. J.* **376**, 404–419 (1991).
33. Schechter, P. *Astrophys. J.* **203**, 297–311 (1976).
34. Salpeter, E. E. *Astr. J.* **106**, 1265–1271 (1993).
35. White, S. D. M., Navarro, J. F., Evrard, A. E. & Frenk, C. S. *Nature* **366**, 429–433 (1993).
36. Mattila, K., Leinert, C. & Schnur, G. in *The Early Observable Universe from Diffuse Backgrounds* (eds Rocca-Volmerange, B., Deharveng, J.-M. & Van, J.-J.-T.) 133–135 (Ed. Frontières, Paris, 1991).
37. Maddox, S. J., Sutherland, W. J., Efsthathiou, G. & Loveday, J. *Mon. Not. R. astr. Soc.* **243**, 692–712 (1990).

ACKNOWLEDGEMENTS. I am grateful to many colleagues for conversations which led to this work, in particular M. Rees, B.E.J. Pagel and S.D.M. White. I acknowledge support from a SERC postdoctoral fellowship.

## Deformation-induced nanocrystal formation in shear bands of amorphous alloys

H. Chen<sup>\*†</sup>, Y. He<sup>†</sup>, G. J. Shiflet<sup>\*</sup> & S. J. Poon<sup>†</sup>

<sup>\*</sup> Department of Materials Science and Engineering,

<sup>†</sup> Department of Physics, University of Virginia, Charlottesville, Virginia 22901, USA

AMORPHOUS alloys formed by rapid solidification of a metallic melt are of considerable technological interest as high-strength materials<sup>1–8</sup>. As they are not in thermodynamic equilibrium, these materials tend to crystallize on heating<sup>9,10</sup>. A high degree of crystallization leads to embrittlement, but if it can be arrested when the crystallites are of only nanometre dimensions, the resulting amorphous–nanocrystalline composite actually has greater strength than the original amorphous material<sup>11</sup>. There is consequently much interest in understanding the mechanisms of crystallization. Previous studies have suggested that mechanical deformation can induce crystallization<sup>12–16</sup>. Here we report the direct observation of crystallization within the shear bands of aluminium-based amorphous alloys induced by bending. The crystals are face-centred cubic aluminium, 7–10 nm in diameter, and seem to form as a consequence of local atomic rearrangements in regions of high plastic strain. We suggest that mechanical deformation might therefore be used to form high-strength amorphous–nanocrystalline composites.

At high temperatures ( $>0.7 T_g$ , where  $T_g$  is the glass transition temperature), amorphous alloys deform homogeneously, and

the flow stress depends strongly on temperature. At low temperatures, they deform elastically at low stress, whereas at high stress levels, inhomogeneous plastic deformation occurs, during which localized narrow shear bands are formed and the plastic flow is confined to the shear band or slipped band region<sup>1</sup>. These shear bands can be produced by compression or tension. Unlike the crystalline materials, there are no well-defined atomic planes nor slip systems in amorphous alloys.

The formation of localized shear bands and the deformation mechanism of amorphous alloys have been the subject of much theoretical discussion<sup>1,3,7</sup> and computer simulations<sup>6,8</sup>. Localization of plastic flow requires that the material within the shear bands undergoes some structural changes, causing a local softening of the material, so the region inside the shear bands deforms more easily than the rest of the sample. The presence of the shear bands reduces the strength of amorphous alloys by providing sites for further plastic flow, as it requires more energy to nucleate a new shear band. Therefore, further deformation develops preferentially along these bands until final fracture occurs<sup>17</sup>. Several early experimental observations suggest that the atomic structure within the shear bands may have changed, because these bands are preferentially etched<sup>14</sup> and exhibit different electron diffraction contrast from the undeformed material<sup>15</sup>.

In the present experiments, we used aluminium-based amorphous alloys obtained by liquid quenching. These amorphous alloys, which were discovered several years ago, contain up to 90 atomic per cent of aluminium and are of the general composition Al–TM–RE, where TM means transition metal, and RE means yttrium and lanthanides<sup>18,22</sup>. These particular metallic glasses have high tensile strengths<sup>18,23</sup> and unusual glass formability<sup>18,23</sup>. The compositions selected were Al<sub>90</sub>Fe<sub>5</sub>Gd<sub>5</sub>, Al<sub>90</sub>Fe<sub>5</sub>Ce<sub>5</sub>, Al<sub>87</sub>Ni<sub>8.7</sub>Y<sub>4.3</sub> and Al<sub>85</sub>Ni<sub>10</sub>Ce<sub>5</sub>. The alloy ingot was prepared by arc-melting nominal amounts of high-purity elements under an argon atmosphere. Metallic glass ribbons 1–2 mm wide, 15–25  $\mu$ m thick and up to several metres long were

<sup>†</sup> Present address: Center for Advanced Materials, Department of Chemistry, University of Massachusetts, Lowell, One University Avenue, Lowell, Massachusetts 01854, USA.



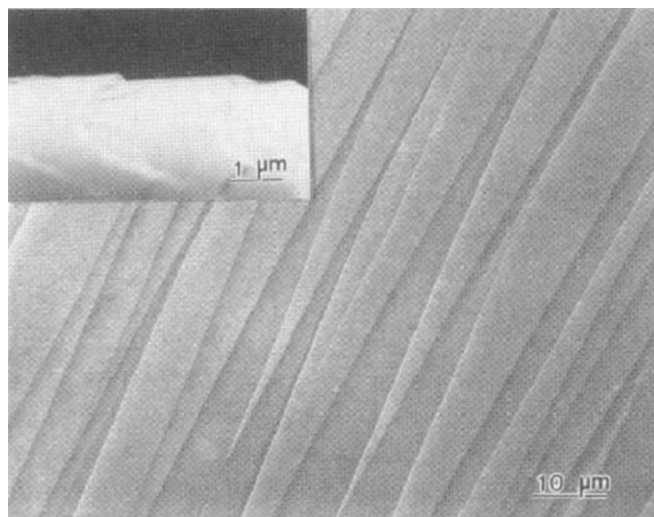


FIG. 1 Main picture, SEM micrograph of the surface of an  $\text{Al}_{90}\text{Fe}_5\text{Gd}_5$  metallic glass ribbon. The sample has been bent at room temperature. The straight lines are slip steps which are terminations of the shear bands on the surface. Inset, an SEM micrograph of the slip steps, taken with electron beam parallel to the bending axis.

obtained by melt-spinning the ingot using a single-roller melt spinner in a partial helium atmosphere. The amorphous nature of the as-spun ribbons was verified using X-ray diffraction, transmission electron microscopy (TEM) and differential scanning calorimetry (DSC) techniques<sup>24</sup>. The ribbons are very flexible and can be bent 180° without cracking. Samples were cut from the as-spun ribbon, and shear bands were introduced by bending the material through 180° at room temperature. The surface of the bent sample was examined using a scanning electron microscope (SEM). TEM specimens were prepared by thinning the bent sample electrolytically in a solution of 25% nitric acid and 75% methanol at 243 K. The microstructure of the bent sample was studied using a 120-keV conventional TEM, and atomic images were obtained employing a top entry high resolution TEM operated at 400 kV.

Examination by SEM reveals a high density of shear bands on the surface of the bent sample near the bending tip (Fig. 1).

These shear bands are imaged as straight lines on the surface and are parallel to the bending axis. The slip steps, of height  $\sim 0.1 \mu\text{m}$ , are shown in the inset. As expected, the density of the shear bands decreases to zero away from the bending tip.

Possible microstructural changes associated with plastic deformation (in this case, bending) were investigated using TEM, focusing on the shear bands themselves and the region between them. Figure 2a shows a TEM micrograph taken near the bending tip of a bent-deformed  $\text{Al}_{90}\text{Fe}_5\text{Gd}_5$  metallic glass ribbon. Running across the micrograph is a shear band (arrowed) which appears lighter in contrast than the rest of the micrograph. Within the shear band, nanometre-size crystalline precipitates are homogeneously distributed. Regions outside the shear band are completely amorphous. Similar phenomena were observed for the bent  $\text{Al}_{90}\text{Fe}_5\text{Ce}_5$  and  $\text{Al}_{87}\text{Ni}_{8.7}\text{Y}_{4.3}$  glassy ribbons. But no such crystalline precipitates were found in the shear bands of bent-deformed  $\text{Al}_{85}\text{Ni}_{10}\text{Ce}_5$  metallic glass, suggesting that there is a compositional-dependence for the mechanical stability against shear deformation-induced crystallization. At lower magnification (Fig. 2b) one can see the extent of waviness in the shear band, but at still lower resolution (Fig. 1) the bands appear straight.

High-resolution TEM reveals (Fig. 3) that the shear band is not completely occupied by nanostructured crystallites, but contains amorphous regions which separate these crystallites. The nano-crystallites, 7–10 nm in diameter, are homogeneously embedded in the amorphous component of the slip band. From the structure image and its corresponding optical diffraction pattern (shown in Fig. 3, inset), the crystal structure of the precipitates is determined to be face-centred cubic (f.c.c.) with a lattice constant of 0.4 nm. Thus these nanoscale crystals are essentially pure aluminium, as the solubility of Fe and Gd in Al is rather small. The structure and morphology of these crystallites produced after bending are the same as that previously observed by us in samples obtained by isothermal annealing<sup>11,24</sup> or by lowering the quenching speed<sup>11</sup>.

To assist further understanding of the influence of the plastic deformation on the structural change of Al-based amorphous alloys, high-energy ball-milling experiments were conducted on these alloys using a standard SPEX-8000 ball mill. The glass ribbons were cut to 5-cm lengths and sealed under argon atmosphere in a hardened tool-steel vial. High hardness, good wear-resistant chrome steel balls (3/8 inch diameter) were used; the ball-to-sample weight ratio was  $\sim 15:1$ . After ball-milling for

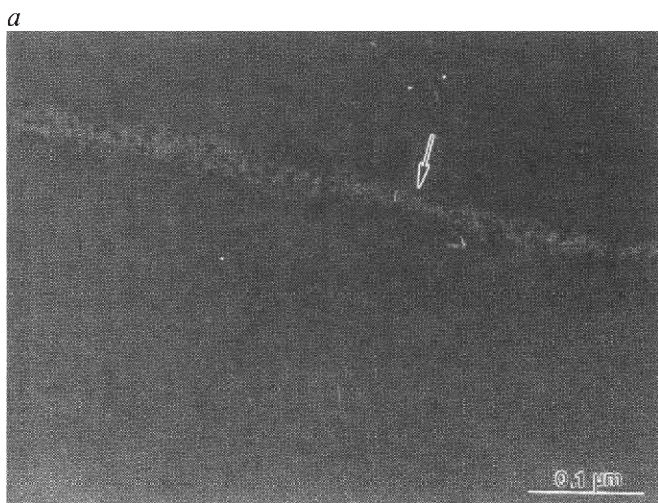
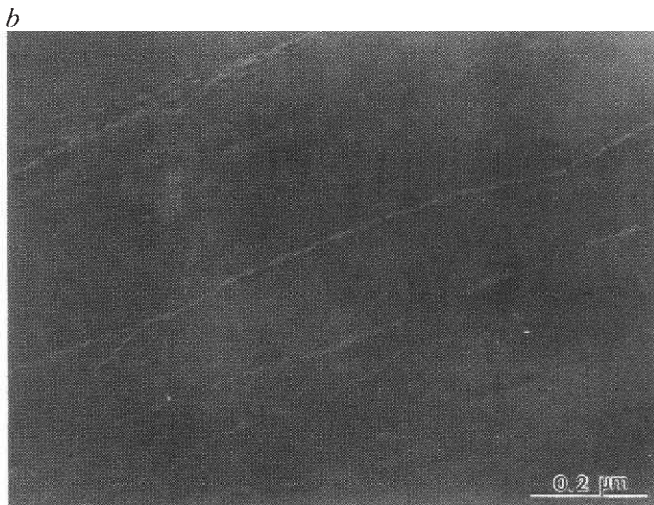


FIG. 2 a, TEM micrograph of a bent  $\text{Al}_{90}\text{Fe}_5\text{Gd}_5$  metallic glass ribbon near its bending tip. A shear band is indicated by an arrow. The white and black spots inside the shear band are the crystalline precipitates, which are surrounded by amorphous matrix. Material outside the shear



band is amorphous. b, TEM micrograph of a bent  $\text{Al}_{90}\text{Fe}_5\text{Gd}_5$  metallic glass ribbon at lower magnification. Several shear bands with nanocrystals may be seen.



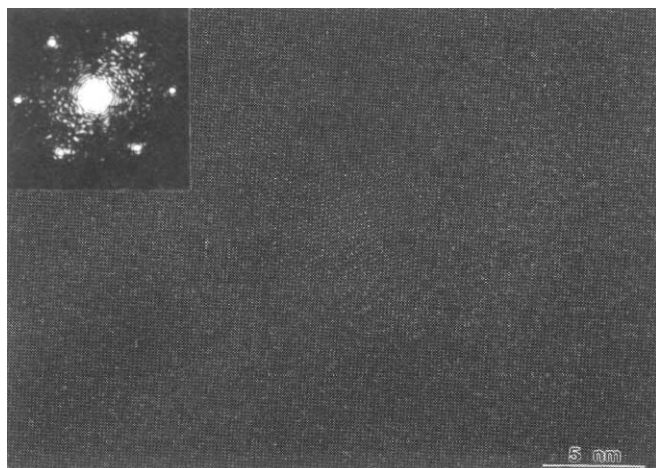


FIG. 3 Main picture, high-resolution TEM atomic image of a nanometre-size crystalline precipitate formed within a shear band of  $\text{Al}_{90}\text{Fe}_5\text{Gd}_5$  metallic glass after bending. The crystallite is embedded in a matrix which remains amorphous. The crystal structure of the precipitate is face-centred cubic with a lattice parameter of 0.40 nm. Inset, the optical Fourier transform of the crystal, showing the orientation to be  $\langle 110 \rangle$ .

1 min, nanocrystals were detected in  $\text{Al}_{90}\text{Fe}_5\text{Gd}_5$  glass by TEM dark-field image technique. After milling for 1 h, both  $\text{Al}_{90}\text{Fe}_5\text{Gd}_5$  and  $\text{Al}_{90}\text{Fe}_5\text{Ce}_5$  showed traces of f.c.c. Al peaks with an amorphous background in the X-ray diffraction pattern, whereas  $\text{Al}_{85}\text{Ni}_{10}\text{Ce}_5$  was still amorphous without detectable crystalline phases in the X-ray diffraction pattern. These results are consistent with the results of structural change by bending. The details of the structural changes induced by ball-milling will be published elsewhere (Y.H., G.J.S and S.J.P., manuscript in preparation).

Previous studies concerned with the influence of plastic deformation on the crystallization behavior of amorphous alloys are summarized in refs 2, 9 and 10. Compression was found to induce crystallization in a glassy  $\text{Fe}_{83}\text{P}_{10}\text{C}_7$  alloy, whereas cold-rolling of amorphous  $\text{Fe}_{80}\text{P}_{13}\text{C}_7$  slowed crystallization<sup>9</sup>. Other reports demonstrated that the crystallization temperature increased with increasing hydrostatic pressure<sup>12,25,26</sup>, which was thought to be due to the reduction of atomic mobility under high pressure. These results indicated that the atomic structure of amorphous alloys may have been changed by plastic deformation. More recently, in the process of studying the tensile deformation of an amorphous  $\text{Al}_{88}\text{Ni}_{10}\text{Ce}_2$  alloy at elevated temperatures, Inoue *et al.*<sup>27</sup> discovered that a large tensile elongation is induced by the formation of nanoscale f.c.c.-Al particles in the sample. This phenomenon was explained as a result of

preferential precipitation of Al particles on the shear plane resulting from the deformation-induced temperature increase during the tensile test. However, these experiments were performed at temperatures sufficiently high to cause crystallization thermally without any external deformation.

One possible explanation of bending-induced crystallization inside the shear bands is the local heating due to local deformation when shear bands are formed. In studying the structural relaxation of a cold-rolled  $\text{Pd}_{77.5}\text{Cu}_6\text{Si}_{16.5}$  amorphous alloy, Chen<sup>28</sup> estimated that the temperature rise at the shear bands is  $>400$  K. For Al-based amorphous alloys, a similar estimate based on adiabatic heating<sup>29</sup> showed that the temperature increase due to local heating may be as high as 2,500 K, which is high enough to induce crystallization. But a simple estimation showed that the heat generated is rapidly conducted away from the shear bands to the rest of the sample on a nanosecond timescale<sup>29</sup>; local heating is therefore unlikely to be inducing crystallization within the shear bands. The experimental result (this work) from glassy  $\text{Al}_{85}\text{Ni}_{10}\text{Ce}_5$  lends support to this argument, as no crystallites were observed inside the slip bands of this material. Thus the observed bending-induced crystallization is probably due to the large permanent strain within the shear bands. Although the atom-scale mechanism of such structural change has not been clarified, it is unlikely that long-range atomic diffusion is involved. On the other hand, the shear strain within the slip bands can be as large as  $10^2$ – $10^3\%$  in the case of bending, and a large proportion of the atoms within the shear bands are subject to local displacements under the stress, as revealed by computer simulation<sup>6</sup>. These local atomic displacements can result in the changes of topological and chemical short-range order of the amorphous state. It is possible that such atomic rearrangements will lead to shifting of Al atoms towards more stable positions and cause the formation of Al nanocrystals, as the environment around an Al atom in the amorphous state is not very different from that in the crystalline state. Once the size of an Al cluster reaches the critical radius  $r_c$  ( $\sim 15$ – $20$  Å), further growth of these Al clusters will be thermodynamically favorable, thus giving rise to the observed nanocrystallites inside the shear bands. Naturally, such shear-strain-induced instability of the amorphous state is highly dependent on the atomic structure and the nature of the chemical bonding in the amorphous alloy.

Our results also show that mechanical treatment provides another means by which nanophase crystallites can be introduced into amorphous alloys. In practice, this could be achieved by pulverization of glassy metals through mechanical alloying; bulk amorphous materials embedded with nanocrystals can be produced by subsequent processing, such as extrusion, warm consolidation and explosive compaction. It is expected that these glassy alloys embedded with nanocrystals will have mechanical properties superior to those of conventional aluminium alloys. □

Received 2 July; accepted 8 December 1993.

- Spaepen, F. *Acta metall.* **25**, 407–415 (1977).
- Masumoto, T. & Maddin, R. *Mater. Sci. Engng* **19**, 1–24 (1979).
- Argon, A. S. *Acta metall.* **27**, 47–58 (1979).
- Kimura, H. & Masumoto, T. in *Amorphous Metallic Alloys* (ed. Luborsky, F. E.) 187–230 (Butterworths, Boston, 1983).
- Spaepen, F. & Taub, A. I. in *Amorphous Metallic Alloys* (ed. Luborsky, F. E.) 231–256 (Butterworths, Boston, 1983).
- Srolovitz, D., Vitek, V. & Egami, T. *Acta metall.* **31**, 335–352 (1983).
- Steif, P. S., Spaepen, F. & Hutchinson, J. W. *Acta metall.* **30**, 447–455 (1982).
- Kobayashi, S., Maeda, K. & Takeuchi, S. *Acta metall.* **28**, 1641–1642 (1980).
- Köster, U. & Herold, U. in *Glassy Metals I* (eds Güntherodt, H. J. & Beck, H.) 225–259 (Springer, New York, 1981).
- Scott, M. G. in *Amorphous Metallic Alloys* (ed. Luborsky, F. E.) 144–168 (Butterworths, Boston, 1983).
- Chen, H., He, Y., Shiflet, G. J. & Poon, S. J. *Scripta Metall. Mater.* **25**, 1421–1424 (1991).
- Iwasaki, H. & Masumoto, T. *J. Mater. Sci.* **13**, 2171–2176 (1978).
- Trudeau, M. L., Schulz, R., Dussault, D. & Van Neste, A. *Phys. Rev. Lett.* **64**, 99–102 (1990).
- Pampillo, C. A. *Scripta Metall.* **6**, 915–918 (1972).

- Donovan, P. E. & Stobbs, W. M. *Acta metall.* **29**, 1419–1436 (1981).
- Walter, J. L., Legrand, D. G. & Luborsky, F. E. *Mater. Sci. Engng* **29**, 161–167 (1977).
- Krishnanand, K. D. & Cahn, R. W. *Scripta Metall.* **9**, 1259–1261 (1975).
- He, Y., Poon, S. J. & Shiflet, G. J. *Science* **241**, 1640–1642 (1988).
- Inoue, A., Ohtera, K., Tsai, A. P. & Masumoto, T. *Jap. J. appl. Phys.* **27**, L479–L482 (1988).
- Shiflet, G. J., He, Y. & Poon, S. J. *Scripta Metall.* **22**, 1661–1664 (1988).
- Inoue, A., Ohtera, K., Zhang, T. & Masumoto, T. *Jap. J. appl. Phys.* **27**, L1583–L1586 (1988).
- Cahn, R. W. *Nature* **341**, 183–184 (1989).
- He, Y., Dougherty, G. M., Shiflet, G. J. & Poon, S. J. *Acta metall. mater.* **41**, 337–343 (1993).
- He, Y., Chen, H., Shiflet, G. J. & Poon, S. J. *Phil. Mag. Lett.* **61**, 297–303 (1990).
- Emmens, W. C., Vrijen, J. & Radelaar, S. J. *non-cryst. Solids* **18**, 299–302 (1975).
- Cedergren, M. & Bäckström, G. J. *non-cryst. Solids* **30**, 69–76 (1978).
- Inoue, A., Kim, Y. H. & Masumoto, T. *Mater. Trans. Jap. Inst. Metals* **33**, 487–490 (1992).
- Chen, H. S. *Appl. Phys. Lett.* **29**, 328–330 (1976).
- Chen, H. *thesis*, Univ. Virginia (1992).

ACKNOWLEDGEMENTS. This work was supported by the US Army Research Office.



# Crossover from creep to inertial motion in friction dynamics

T. Baumberger, F. Heslot & B. Perrin

Laboratoire de Physique de la Matière Condensée, Laboratoire Associé au CNRS et aux Universités Paris 6 et Paris 7, Ecole Normale Supérieure, 24 rue Lhomond, 75231 Paris Cedex 05, France

Friction between dry surfaces plays a role in solid dynamics over a wide range of length scales, from the mechanics of micro-machines to earthquakes. Some common features of frictional dynamics have been observed for rather different materials, such as rock sliding on rock or metal on metal<sup>1,2</sup>. But there is still relatively little understanding of the way in which frictional motion varies generically with mechanical parameters. Here we present the results of experiments on frictional sliding of Bristol board, which show how the characteristics of frictional sliding depend on mass, driving velocity and stiffness of the driving spring constant. In general, there is a bifurcation from stick-slip to steady sliding with increasing velocity. The character of the frictional motion and of the associated bifurcation changes with velocity: at low speeds creep is dominant, which can be ascribed a characteristic length, whereas at high speeds the motion can be described as inertial, with a characteristic timescale. The kinetic friction coefficient decreases with increasing velocity in the former case, and increases with velocity in the latter. We anticipate that these results are likely to be generic.

When two solids in contact are rubbed together without lubrication, their relative motion usually proceeds with jerky oscillations rather than in a smooth fashion<sup>3</sup>. Driven at constant speed, these oscillations essentially involve a 'stick' state associated with the elastic loading of the system, and a sudden 'slip' corresponding to stress relaxation. The standard mechanism for 'stick-slip' involves a static threshold (force necessary to start motion), larger than the force of friction during sliding<sup>1</sup>. Stick-slip occurs

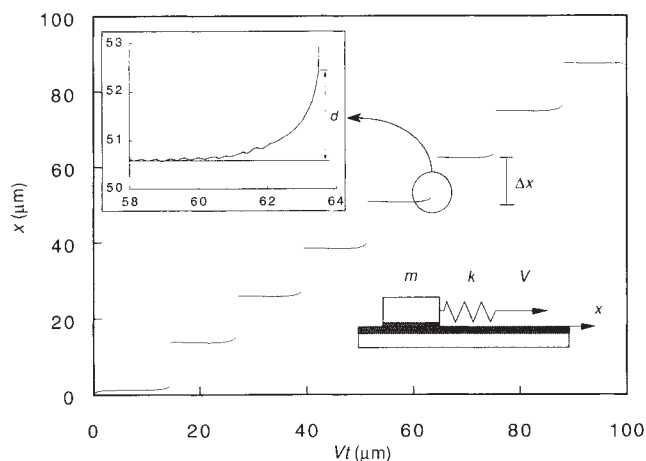


FIG. 1 Time ( $t$ ) evolution of the slider displacement ( $x$ ) with respect to the bench, in the stick-slip phase. A single event is magnified (inset) to show the creep precursor to the actual slip. The creep distance  $d$  and the slip distance  $\Delta x$  are shown. The apparatus is sketched in the main figure. The surfaces in contact are made of Bristol board, the nominal contact area being  $8 \times 9 \text{ cm}^2$ . The slider of mass  $m$  is pulled by a spring of stiffness  $k$  (cylindrical or cantilever, depending on range) driven at constant velocity  $V$ . The parameters of the main figure are  $k = 580 \text{ N cm}^{-1}$ ,  $m = 800 \text{ g}$  and  $V = 1 \mu\text{m s}^{-1}$ . The instantaneous spring elongation is measured using a magnetic displacement transducer.

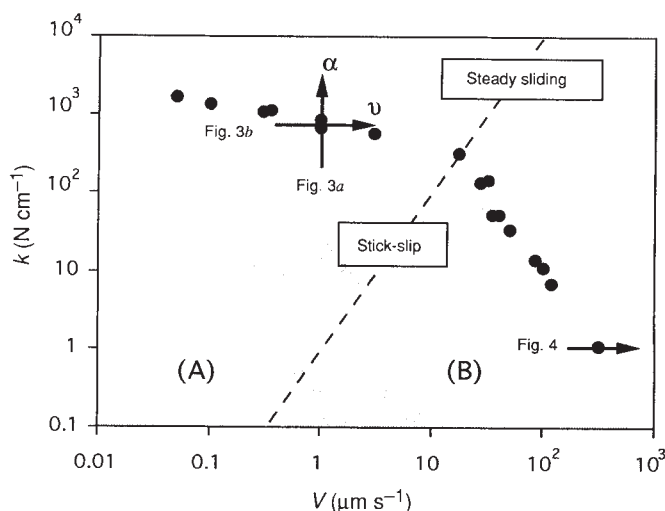


FIG. 2 Plot (dots) in the  $k-V$  plane of the bifurcation curve  $k_c(V)$ . Sliding is stable above the curve (unshaded area) whereas stick-slip oscillations occur below the curve (shaded area). As the data range over decades, a log-log plot is used for clarity. Taking into account the frictional behaviour, as revealed by the experiments, two regions in the  $k-V$  plane may be distinguished: the creep region (A) and the inertial region (B). The dashed line, separating both regions and crossing the bifurcation curve at its shoulder, corresponds to  $T_{\text{creep}} = 1.5 \times T_{\text{inertial}}$ , that is  $k = (1.5)^2 (2\pi)^2 / d^2 m V^2$ . It matches the change in the nature of the bifurcation: supercritical Hopf in (A) where a control parameter is  $k/m$ , and subcritical in (B) where the scaling  $k/m$  is less clear. In (B), data indicate the onset of intermittency and are obtained with  $m = 1.2 \text{ kg}$ ; in (A), data obtained with  $m$  ranging between  $300 \text{ g}$  and  $3 \text{ kg}$  are scaled to  $m = 1.2 \text{ kg}$ . Arrows labelled Fig. 3a, Fig. 3b and Fig. 4 indicate regions that are examined in more detail in the relevant figures:  $\alpha$  and  $v$  are the associated control parameters.

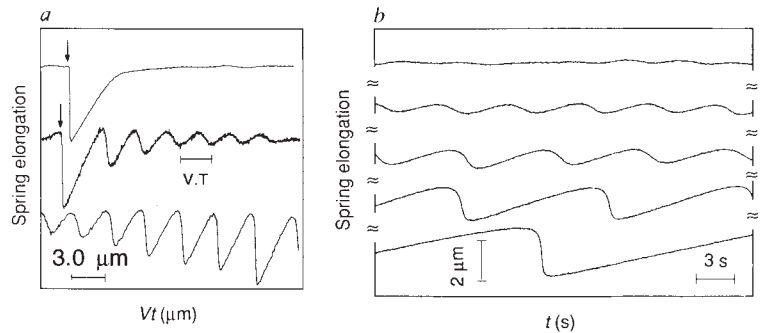
in many systems<sup>1-6</sup> and has been claimed to be a basic component of earthquakes<sup>7</sup>.

To investigate basic features of friction dynamics, we have used an experimental model system. It consists of a rigid slider of mass  $m$  driven (on a bench) at constant speed  $V$  by means of a spring of stiffness  $k$  (Fig. 1). The mechanical parameters are  $V$ ,  $k$  and  $m$ , and range over decades, namely  $V$  from  $10^{-2} \mu\text{m s}^{-1}$  to  $5 \text{ cm s}^{-1}$ ,  $k$  from  $1 \text{ N cm}^{-1}$  to  $10^4 \text{ N cm}^{-1}$ , and  $m$  from  $300 \text{ g}$  to  $3 \text{ kg}$ . Both surfaces in contact are made of Bristol board, which is composed of isotropically agglomerated polymeric fibers ( $\sim 10 \mu\text{m}$  wide). Stick-slip oscillations (Fig. 1) are favoured by low  $k$ , high  $m$  and low  $V$ , and have typical slip distances  $\Delta x$  (Fig. 1) ranging from  $1 \mu\text{m}$  to  $10 \text{ cm}$ . They are remarkably periodic and reproducible, although for a given sample a slow drift of the friction properties is observed over weeks, attributed to wear of the contact surfaces.

We have mapped an extensive 'phase diagram' (in the  $k-V$  plane) of friction dynamics. Stick-slip or steady sliding occur, respectively below and above a bifurcation curve  $k_c(V)$  (Fig. 2). As seen from the marked change of slope, two different regimes may already be suspected, called here (for reasons explained later) the 'creep' regime (A), corresponding to low speed, and the 'inertial' regime (B) at high speed.

We first address the creep regime. At low velocity, it appears from Fig. 1 that just before rapid slip, the loading signal shows a marked rise. Actually, there is no true stick phase and the elastic stress partially relaxes through a creep process. As already identified at low speed in rock-rock<sup>2</sup> and metal-metal<sup>1</sup> friction, this creep motion occurs over a characteristic length  $d$  of the order of micrometres, independent of  $V$ . Moreover, here we characterize experimentally the nature of the bifurcation: (1) Approaching the bifurcation line, for example by increasing  $V$ ,

FIG. 3 Dynamical characteristics of the system, close to the bifurcation, in the creep regime. The spring elongation is represented. An offset is added to each curve for sake of clarity. *a*, Transient responses of the system at constant  $V=1\text{ }\mu\text{m s}^{-1}$  but at various values of the reduced parameter  $\alpha=k/k_c-1$  (see Fig. 2). From lower to upper curve:  $\alpha=-0.1$ ;  $0.4$ ;  $>2$ . The two upper curves correspond to a speed impulse perturbation, indicated by arrows; the lower curve is obtained after quenching the system in the stick-slip phase, by increasing  $m$ . The pseudo-distance  $VT$  defined in the text is shown. A similar evolution of relaxation has been observed in a rock-rock system, (A. Ruina, personal communication). *b*, Bifurcation sequence for the steady-state amplitude of oscillations in the stick-slip phase, following here a path at constant  $k$  and  $m$  (see Fig. 2). The distance to the bifurcation line is given by the reduced velocity  $v=(V-V_c)/(V_c)$ , where  $V_c(k/m)$  is the bifurcation speed. From lower to upper curve:  $v=-0.75$ ;



$-0.58$ ;  $-0.41$ ;  $-0.25$ ;  $\sim 0$ . Note that the inertial time, here  $T_{\text{inertial}}=24\text{ ms}$ , is much smaller than the timescale involved in this figure.

the stationary amplitude of stick-slip oscillations tends continuously to zero (Fig. 3*b*). It is also possible to drive the system across the bifurcation by varying the stiffness  $k$  or the normal load (here  $mg$ , where  $g$  is the acceleration due to gravity); we find that  $k/m$  is the relevant control parameter in this regime. Qualitatively, stick-slip becomes nearly sinusoidal when the slip distance  $\Delta x$  is of the order of the creep distance  $d$ , that is for  $k=(mg)\times\Delta\mu/d$ , with  $\Delta\mu\approx 10^{-2}$ , representing a typical friction drop during a slip. (2) Speed impulse perturbations in the steady sliding phase give the transient characteristics of the system (Fig. 3*a*): starting deep in the steady sliding zone and heading towards the transition line, the response goes from aperiodic to periodic (of period  $T$ ) with a diverging damping time. Ultimately, quenching the system in the stick-slip region, periodic amplified oscillations saturate to stick-slip.

The characteristics (1) and (2) suggest that this transition is a supercritical Hopf bifurcation<sup>8</sup>. The measured quantity  $VT$  is roughly a constant along the bifurcation line and again defines a length (typically  $2\text{--}4\text{ }\mu\text{m}$ ), comparable to the creep distance. Characteristic distance in creep and transients has been reported in rock friction experiments, and is one of the basic features of phenomenological models<sup>9–12</sup>.

We address now the high-speed regime (region B in Fig. 1). In contrast to the creep bifurcation, the bifurcation from stick-

slip to steady sliding occurs at finite stick-slip amplitude, much larger than  $d$  (Fig. 4). Even near the transition, the slip is abrupt, and occurs over the time  $\pi(m/k)^{1/2}=T_{\text{inertial}}/2$  (half-eigenperiod of the system). Close to the bifurcation, the system is very sensitive to noise, either intrinsic (spatial inhomogeneities of the track) or extrinsic (external vibration). Noise may lead to temporal intermittency between stick-slip and smooth sliding (Fig. 4). If the noise level is reduced, a hysteresis loop between stick-slip and steady sliding is observed, confirming the marked subcritical nature of the high-speed bifurcation<sup>8</sup>. Within the loop, the system can be brought either into the stick-slip or the smooth sliding regime by *ad hoc* finite-amplitude perturbations. The data shown in Fig. 2 correspond to intermittent behaviour. When perturbed in the steady sliding zone, the system responds through a classical free damped oscillation, of period  $T\approx T_{\text{inertial}}$ .

We can formulate a criterion to separate the two regimes: in (A), the time of creep over the system characteristic length  $d$  is  $T_{\text{creep}}\approx d/V$  and dominates the dynamics, whereas in (B),  $T_{\text{inertial}}$  becomes the (slower) dominating time. The change of dynamical regime corresponds to  $T_{\text{creep}}$  of order  $T_{\text{inertial}}$  (see the dashed line in Fig. 2).

The change of regime between creep and inertial motion is moreover accompanied by changes in the velocity-dependence of steady friction coefficients. The steady-state kinetic friction coefficient  $\mu_{\text{kinetic}}^{\text{steady}}(V)$  is determined in the steady regime, as the tangential force normalized to the weight. In the creep regime,  $\mu_{\text{kinetic}}^{\text{steady}}(V)$  is velocity-weakening (decreasing with velocity); for  $V$  in the range  $0.1\text{ }\mu\text{m s}^{-1}$  to  $100\text{ }\mu\text{m s}^{-1}$  we can use the fit  $\mu_{\text{kinetic}}^{\text{steady}}(V)=\mu_k^1-\delta\log(V)$ , with  $\mu_k^1=0.37$  and  $\delta=3\times 10^{-2}$  (such a log-dependence has been reported for rock-rock data<sup>2</sup>). We emphasize that this measurement is performed on a steady motion at constant velocity and cannot be extrapolated to non-stationary motion: indeed, assuming a velocity weakening for the instantaneous velocity would lead to linear instability of any perturbation, incompatible with the observed stationary motion. This point is usually taken into account in phenomenological models, through a 'memory length'  $D_0$  of order of the creep distance  $d$ , associated with a history-dependent friction coefficient. Following various authors<sup>1,9,10</sup>, it is tempting to connect  $D_0$  with a characteristic of the surfaces in contact, for example, the mean displacement needed to renew the population of contact spots (when no gouge is present<sup>13</sup>). By contrast, in the inertial regime,  $\mu_{\text{kinetic}}^{\text{steady}}(V)$  is velocity-strengthening (increasing with velocity), approximated by  $\mu_{\text{kinetic}}^{\text{steady}}(V)=\mu_k^0+\eta V$ , with typically  $\mu_k^0\approx 0.3$  and  $\eta\approx 3\text{ s m}^{-1}$ . The velocity  $V^*$ , at which the slope of  $\mu_{\text{kinetic}}^{\text{steady}}(V)$  changes sign, is  $k$ - and  $m$ -dependent. More precisely, when  $k/m$  is varied,  $V^*$  is shifted along the separation creep/inertia (dashed line in Fig. 2). So we stress that the onset of the inertial regime and the associated velocity strengthening behaviour is system-dependent. As  $T_{\text{creep}}$  is here much smaller than  $T_{\text{inertial}}$ , memory effects are negligible. We may thus assume

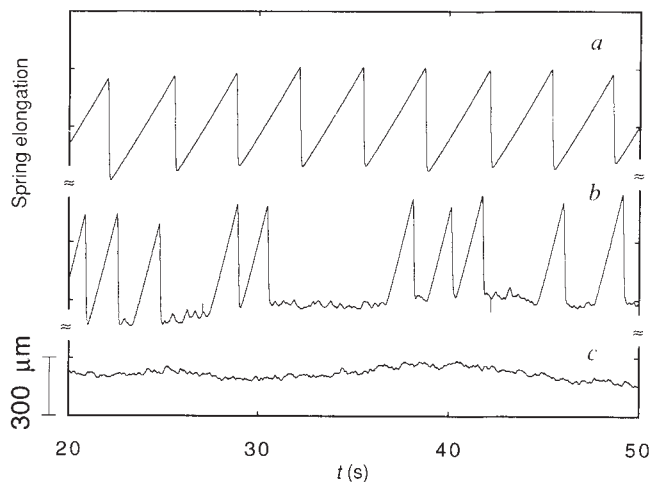


FIG. 4 Bifurcation sequence in the inertial regime. The transition to steady sliding occurs by way of temporal intermittency. Driving velocities  $V$  are  $275\text{ }\mu\text{m s}^{-1}$  (a),  $525\text{ }\mu\text{m s}^{-1}$  (b), and  $890\text{ }\mu\text{m s}^{-1}$  (c), with  $k=1\text{ N cm}^{-1}$  and  $m=1.2\text{ kg}$ . Note that the amplitude of the slip events at the bifurcation is much larger than the creep distance  $d$ .



that the expression for  $\mu_{\text{kinetic}}^{\text{steady}}(V)$  can be extended to instantaneous sliding velocity. Velocity strengthening then leads to damping. It implies linear stability of steady sliding but does not prevent finite amplitude instability (stick-slip). Physically, damping may be related to vibrations induced by collisions between contact spots. Such a feature has been introduced in earthquake models to account for radiative elastic dissipation (see ref. 12 for example).

The low-speed regime of this study, together with standard results of rock-rock and metal-metal friction, provides strong support for the universality of low-velocity friction dynamics, which appears quite independent of the structure and nature of the underlying materials. Moreover, the criterion for the crossover to the high-speed regime being extremely simple, one may expect the inertial regime encountered in this study to also be of general relevance. Recent attempts to extend the range of validity of previous models<sup>11</sup> may include, as a phenomenological parameter, the crossover to the high-speed transition

reported here. A connection between this crossover and a microscopic length is provided here as an attempt to elucidate further the physical processes behind friction dynamics. □

Received 1 June; accepted 9 December 1993.

1. Rabinowicz, E. *Friction and Wear of Materials* (Wiley, New York, 1965).
2. Scholz, C. H. *The Mechanics of Earthquakes and Faulting*, Ch 2 (Cambridge Univ. Press, 1990).
3. Bowden, F. P. & Tabor, D. *The Friction and Lubrication of Solids* Ch 5 (Clarendon, Oxford, 1950).
4. Oden, J. T. & Martins, J. A. C. *Comp. Math. appl. mech. Engng.* **52**, 527–634 (1985).
5. Jeager, H. M. & Nagel, S. R. *Science* **255**, 1523–1531 (1992).
6. Yoshizawa, H., McGuigan, P. & Israelachvili, J. *Science* **259**, 1305–1308 (1993).
7. Brace, W. F. & Byerlee, J. D. *Science* **15**, 990–992 (1966).
8. Jackson, E. A. *Perspectives of Nonlinear Dynamics* (Cambridge Univ. Press, 1986).
9. Dieterich, J. H. *J. geophys. Res.* **84**, 2161–2168 (1979).
10. Rice, J. R. & Ruina, A. L. *J. appl. Mech.* **50**, 343–349 (1983).
11. Gu, Y. & Wong, T.-F. in *Rock Mechanics Proc. 33rd US Symp.* (eds Tillerson, J. R. and Wawersik, W. R.) 151–158 (Balkema, Rotterdam, 1992).
12. Rice, J. R. *J. geophys. Res.* **98**, 9885–9907 (1993).
13. Marone, C. & Kilgore, B. *Nature* **362**, 618–621 (1993).

## In situ speciation measurements of trace components in natural waters using thin-film gels

W. Davison & H. Zhang

Institute of Environmental and Biological Sciences,  
Environmental Science Division, Lancaster University,  
Lancaster LA1 4YQ, UK

RELIABLE measurement of trace species in natural waters is essential for studies of pollution or trace-element cycling, but is difficult, partly because the distribution of chemical species often changes during sampling and storage<sup>1</sup>. *In situ* measurements can overcome these problems, but the few measurements made previously have involved complicated systems that cannot be used routinely<sup>1,2</sup>. Here we describe a simple technique for measuring trace-metal concentrations *in situ* in water. The technique incorporates an ion-exchange resin separated from the solution by an ion-permeable gel membrane. Mass transport through the gel is diffusion-controlled and thus well defined, making it possible to obtain quantitative data on concentration and speciation over relatively short time periods (from one hour to several weeks). We present measurements of zinc concentrations in sea water using this technique which agree well with electrochemical measurements. In principle, our technique should be applicable to any inorganic or organic diffusing species.

*In situ* measurement of trace components in natural waters has proved an elusive goal. True equilibrium procedures, such as ion-selective electrodes or dialysis, suffer from poor sensitivity and selectivity or unrealistic equilibration times. Other speciation procedures, including anodic stripping voltammetry (in which metals are accumulated at an electrode before analysis) and the use of ion-exchange resins, are kinetically controlled and so the measured species depends on the chemical reaction involved and the rate of mass transport of ions from the bulk solution to the reaction site<sup>3</sup>. To provide quantitative measurements the mass transport must be controlled. Voltammetric measurements provide this control by relying on very short time measurements, externally induced convection or more recently on microelectrodes where diffusion processes dominate. Their capabilities for *in situ* measurements have been extensively

exploited for the measurement of oxygen, but reports of measurements of metals, which require relatively sophisticated on-site equipment<sup>3</sup>, are so far very limited<sup>1</sup>. Ion-exchange resins and organic absorbants can be used for a wider range of components. But if they are simply suspended in bags in natural waters the rate of transport of ions from solution to the reaction site varies with time due to irregular convective processes, resulting in semi-quantitative data<sup>4</sup>. Here we introduce a new simple way of controlling mass transport which enables quantitative data on concentration and speciation to be obtained using any medium capable of reacting with the species of interest. The idea is simply to introduce a known thickness of gel between the solution and the reactive medium. Transport in the gel is restricted to diffusion and, by selection of an appropriate gel thickness, it controls the overall rate of mass transport irrespective of the hydrodynamics in the bulk solution. The procedure has provided *in situ* speciation measurements using simple, inexpensive and readily available equipment which are capable of theoretically sound interpretation.

Polyacrylamide gels have been used to provide measurements of pore waters by a new technique of diffusive equilibration in

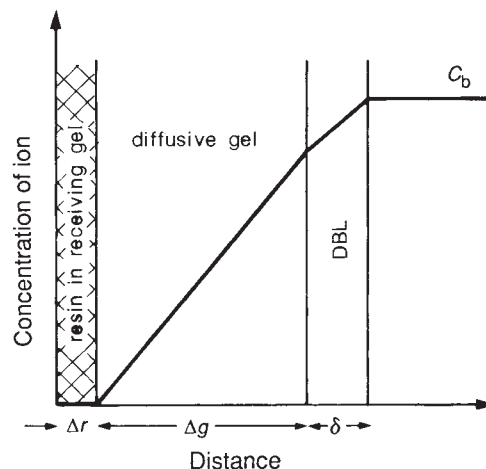


FIG. 1 Schematic representation of the free concentration of ionic species in a gel assembly in contact with natural water where the concentration is  $C_b$  (DBL, diffusive boundary layer.)

thin films (DET)<sup>5</sup>. Ions are simply allowed to diffuse into the gel until equilibrium with the pore waters is established. Here we have used a similar polyacrylamide gel, but have backed it by a further thin film ( $\sim 150\text{ }\mu\text{m}$  thick) of gel containing a cation-exchange resin selective for trace metals (Chelex 100)<sup>6</sup> close packed in a single layer of  $75\text{--}150\text{ }\mu\text{m}$  spheres (Fig. 1). Within the layer of resin, of thickness  $\Delta r$ , the concentration of the free metal in solution is effectively zero, owing to the complexation by the resin which has been reacted to only a small fraction of its capacity. Within the bulk solution the metal has concentration  $C_b$ . The gel layer is assumed to be separated from the bulk solution by a diffusive boundary layer (DBL), of thickness  $\delta$ , where transport is solely by molecular diffusion<sup>7</sup>. To be transported from the solution to the resin, ions must diffuse across the DBL and then through the gel, of thickness  $\Delta g$ . Small ions can diffuse freely through the effectively  $2\text{--}5\text{ nm}$  pores<sup>8</sup> of polyacrylamide gel with effective diffusion coefficients,  $D$ , indistinguishable from those in water<sup>9</sup> (W.D. and H.Z., unpublished results). If for simplicity, it is assumed that  $\delta \ll \Delta g$ , Fick's laws can be used to define the flux of a given metal ion

$$\text{flux} = DC_b/\Delta g \quad (1)$$

The mass per unit area of resin,  $M_a$ , after time  $t$  is then:

$$M_a = DC_b t/\Delta g \quad (2)$$

and the concentration in the resin layer,  $C_r$ , is given by:

$$C_r = M_a/\Delta r \quad (3)$$

After a given time, the concentration in the resin layer,  $C_r$ , can be measured and the concentration in the solution can be quantified by:

$$C_b = C_r \Delta g \Delta r / Dt \quad (4)$$

Providing the resin is not saturated, the longer such a device is immersed, the more metal will be accumulated and the ratio of the concentration of metal in the resin layer to metal in solution will increase as the thickness of the resin and gel layers are decreased. Assuming a typical value of  $D$  of  $10^{-5}\text{ cm}^2\text{ s}^{-1}$ , for a 24-h immersion, a gel layer thickness of 1 mm and a resin layer thickness of 0.1 mm, the concentration in the resin layer will be 864 times greater than the concentration in the bulk solution. Such a procedure therefore offers a large concentration enhancement for relatively short immersion times. This arrangement of a resin backing a gel relies on measuring a flux of metal over a given time. It depends on establishing a defined diffusion gradient in a thin film (DGT), in contrast to DET which depends on equilibrium being established. Equations (1)–(4) depend on the assumption that the thickness of the DBL between the gel and the bulk solution is negligibly small. The only available information on DBL thicknesses in natural waters applies to the sediment–water interface. In the bottom waters of stratified lakes or deep-sea locations, estimates of the DBL thickness are typically  $\sim 1\text{ mm}$  (ref. 7). Although definitive data are not available, various estimates suggest a range of  $0.1\text{--}0.01\text{ mm}$  (ref. 10) may be appropriate in faster-moving waters, such as rivers and the surface waters of lakes and seas. If the gel layer is 1 mm thick, variation in  $\delta$  between 0.1 and 0.01 mm could at most result in a change in flux to the Chelex of 10%. By ensuring that the gel layer is sufficiently thick, the DGT technique can in principle control the mass transfer of metal ions irrespective of changes in the velocity of water in the bulk solution.

The species measured by DGT can be appreciated by considering a simple equilibrium between free metal,  $M$ , and ligand,  $L$ . Binding between  $M$  and the resin effectively maintains  $M$  at a much lower concentration in the resin layer than in the bulk solution. If  $M$  binds more strongly to the resin than the ligand,  $L$ , and the release of  $M$  from  $ML$  is a fast process, there will effectively be a zero concentration of  $M$  and  $ML$  in the resin layer. Thus, as well as there being a concentration gradient for

$M$  between the resin layer and the bulk solution (Fig. 1), there will be a concentration gradient for  $ML$ . If there is a very fast equilibrium between  $M$  and  $ML$ , they will both effectively contribute to the accumulating metal within the resin layer. For a very slow equilibrium only  $M$  will contribute. The physical gel layer between the resin and solution acts as a reaction layer which defines the measured species kinetically in an analogous fashion to an anodic stripping experiment, where the theory is already established<sup>11</sup>. The thickness of the gel layer determines an effective time of measurement. At 1 mm thickness this characteristic time is  $\sim 5\text{ min}$ , longer than the  $0.1\text{--}1.0\text{ s}$  typical of voltammetric experiments. DGT can therefore be expected to measure those species which (1) can readily dissociate in 5 min and (2) have a stability constant effectively smaller than that characterizing the binding of the metal to resin. The gel matrix can also be expected to exclude very large molecules. Although the above theory

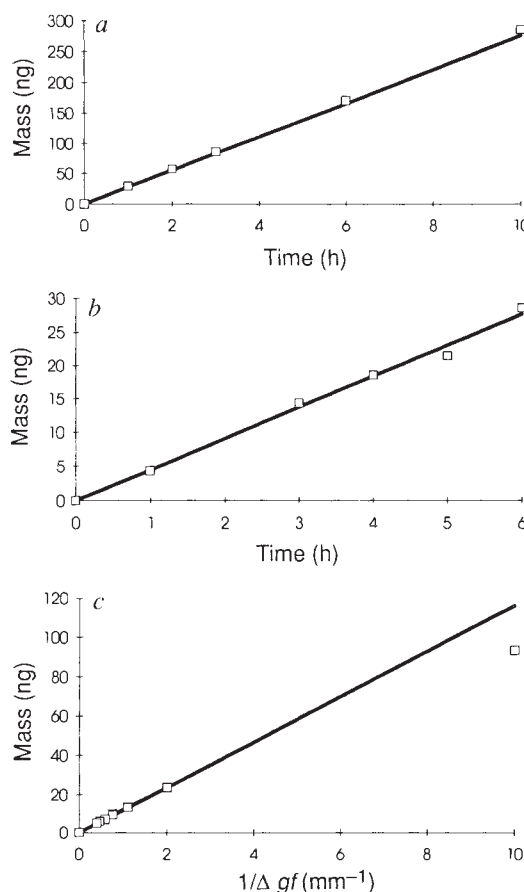


FIG. 2 a, Measured mass of zinc in the receiving resin (hollow squares) for gel assemblies suspended for different times in a stirred solution of natural sea water (pH 7.8) in the laboratory ( $28\text{ }^{\circ}\text{C}$ ). The concentration of labile zinc was measured by direct anodic stripping voltammetry to be  $31\text{ nM}$  and was used with equation (2) to calculate the straight line. The good fit with the experimental data demonstrates the excellent agreement between anodic stripping voltammetry and the diffusion gradient in a thin film (DGT) technique in measuring labile zinc in a natural sea water sample. b, Measured mass of zinc in the receiving resin (hollow squares) for gel assemblies with a  $0.4\text{-mm}$  gel layer suspended in sea water (Menai Straits, UK, salinity  $32\text{‰}$ ,  $14\text{ }^{\circ}\text{C}$ ) for various times. To prevent accumulation of particulates, the assemblies were covered with a  $100\text{-}\mu\text{m}$  thick,  $0.45\text{-}\mu\text{m}$  pore size Millipore cellulose nitrate membrane. Concentrations estimated without the membrane were indistinguishable. c, Measured mass of zinc in the receiving resin for gel assemblies of various combined gel and filter layer thickness,  $\Delta g$ , exposed to sea water (as above) for  $5\text{ h } 20\text{ min}$ . An inverse relationship is predicted by equation (2).



requires detailed validation, for example it may be compromised by metal fulvic complexes adsorbing directly to the resin, it is clear that this new speciation measurement is not dependent on the nature of the resin providing the reaction between free metal and resin is fast. Other sorbent or immobilized ligands, which could compete effectively with the solution ligands and would not adsorb metal complexes, could be used and the same species would be measured.

To test these principles, gel assemblies were constructed using a 10.5 cm diameter, 1.3 cm total thickness perspex disc containing layers of gel and resin (Fig. 1). The gel was exposed through a 5.0 cm diameter window; a set of screws and an 'o'-ring prevented ingress of water to the side or back of the gel while holding the layers of gel firmly onto the perspex backing. After casting, polyacrylamide gel was hydrated in water for at least 24 h to ensure dimensional stability before use. The resin was embedded in a separate  $\sim 150\text{ }\mu\text{m}$ -thick gel as a single plane of approximately close-packed beads. After immersion, the gel layer was peeled off, metal extracted from the resin layer with 1 ml of 2 M  $\text{HNO}_3$  and measured using atomic absorption spectroscopy. Self-diffusion coefficients for Zn for the appropriate temperatures<sup>9</sup> were used. Gel thicknesses should be entirely reproducible as they are determined by casting. The thickness of the hydrated gel was measured by a travelling microscope and was accurate to better than 4%.

Laboratory exposures of gel assemblies to stirred solutions of  $\sim 10^{-7}\text{ M Zn(NO}_3)_2$ , with and without added NaCl (0.5 M), showed that the concentration of metal measured in the resin layer could be predicted quantitatively (97–100%) by equation (4). Measurements of Zn in pH-adjusted sea-water, covering the pH range 4.8–7.8, by direct anodic stripping voltammetry without sample pretreatment and by gels exposed to stirred solutions were in good agreement (Fig. 2a). In an unstirred solution recovery was only 60%, indicating that mass transport was partly limited by the DBL in solution. On lowering the pH to 2.5, recovery was greatly reduced, consistent with the properties of Chexel 100.

In natural sea water, particles adhere to the gel surface and so may affect the inward diffusion of metal ions. Covering the gel with a 100  $\mu\text{m}$ -thick, 0.45- $\mu\text{m}$  pore size cellulose nitrate membrane prevented the particles sticking. Diffusion through the filter was indistinguishable from that through a similar thickness gel. Varying the time of *in situ* immersion in sea water (Fig. 2b) resulted in mass increasing linearly with time. As the tidal current varied from 0 to  $\sim 4$  knots during this time, the thickness of the DBL will also vary. Consequently the linear response indicates that the gel thickness is dominating the control of mass transport confirming the assumption that the DBL is negligible. The mean concentration from these measurements was very reproducible at  $11.9 \pm 0.4\text{ nM}$  as compared to  $26.5 \pm 2.8\text{ nM}$  from seven samples taken at hourly intervals and measured by anodic stripping voltammetry after acidification to pH 2 and exposure to ultraviolet irradiation. A difference is to be expected if it is considered that DGT only measures labile species and therefore will exclude kinetically inert organic species and large colloids.

To investigate further the effect of the gel layer, assemblies of different gel thicknesses, covered by 100  $\mu\text{m}$ -thick filters, were exposed to sea water for 320 min. Consistent with equation (2), the mass measured in the resin was inversely proportional to the total thickness of the filter and gel layer (Fig. 2c), except for when there was no gel layer and only a 100- $\mu\text{m}$  filter was used. The reduced recovery for the 100- $\mu\text{m}$  case can be used to provide an estimate of the mean effective thickness of the DBL of 30  $\mu\text{m}$ .

These measurements demonstrate that the theoretical principles of the DGT technique hold for practical *in situ* use. By using a gel to define the diffusion layer thickness, the mass transport of the system and the criteria for defining the measured species are effectively defined. Using a 0.5-mm diffusion layer, as in these experiments, a concentration factor of 72 is obtained by a 1 h immersion. Such a procedure using simple equipment

immediately overcomes most contamination problems that beset trace metal measurements. The technique is suitable for deployment from aboard ship and a time of 1 h is appropriate for sensing temporal changes in trace metal concentrations in sea water. Changes during tidal cycles could potentially be measured, especially as the technique is independent of ionic strength and pH. Moreover, as the capacity of the resin should allow 3 months immersion in relatively contaminated coastal waters before saturation is reached, the device can readily be used to provide a long-term (weeks, months) integrated record of trace metal concentrations. The limiting factor will most likely be the effects of biofouling, but there is certainly potential for the use of DGT as a chemical alternative to shellfish for providing integrated records of trace metal pollution<sup>12</sup>. DGT is in principle applicable to most trace metals in any non-acid aqueous medium, although the detailed use in complicated solutions like fresh waters, where colloidal material is very important, requires further testing. This new measurement principle can theoretically be applied to any component, including organics, that can readily diffuse through a gel layer and be consumed by an active component in a backing layer. □

Received 19 July; accepted 22 December 1993.

1. Tercier, M. L., Buffle, J., Zirino, A. & De Vitre, R. R. *Analytica chim. Acta* **237**, 429–437 (1990).
2. Davey, E. W. & Soper, A. E. *Limnol. Oceanogr.* **20**, 1019–1023 (1975).
3. Buffle, J. *Complexation Reactions in Aquatic Systems* (Ellis-Horwood, Chichester, 1988).
4. Morrison, G. M. P. *Envir. Technol. Lett.* **8**, 393–402 (1987).
5. Davison, W., Grime, G. W., Morgan, J. A. W. & Clarke, K. *Nature* **352**, 323–325 (1991).
6. W. E. Reinlove. *J. phys. Chem.* **96**, 10–19 (1993).
7. Santschi, P., Hohener, P., Benoit, G. & Buchholtz-ten Brink, M. *Mar. Chem.* **30**, 269–315 (1990).
8. Chrambach, A. *The Practice of Quantitative Gel Electrophoresis* (VCH, Weinheim, 1985).
9. Li, Y. & Gregory, S. *Geochim. cosmochim. Acta* **38**, 703–714 (1974).
10. Snodgrass, W. J. in *Sediments and Water Interactions* (ed. Sly, P. G.) (Springer, New York, 1986).
11. Davison, W. *J. Electroanal. Chem.* **87**, 395–404 (1978).
12. Lopez, P. B., Bajdik, C. D., Belkhole, S. P., Jackson, S. E. & Longerich, H. P. *Arch. Envir. Contam. Tox.* **21**, 409–420 (1991).

ACKNOWLEDGEMENTS. We thank K. Reid, G. Johnson and G. Jones for help with sampling, and the UK Natural Environment Research Council for financial support.

## Western boundary currents in the atmosphere of Mars

M. M. Joshi, S. R. Lewis, P. L. Read & D. C. Catling

Atmospheric, Oceanic and Planetary Physics, Clarendon Laboratory, Parks Road, Oxford OX1 3PU, UK

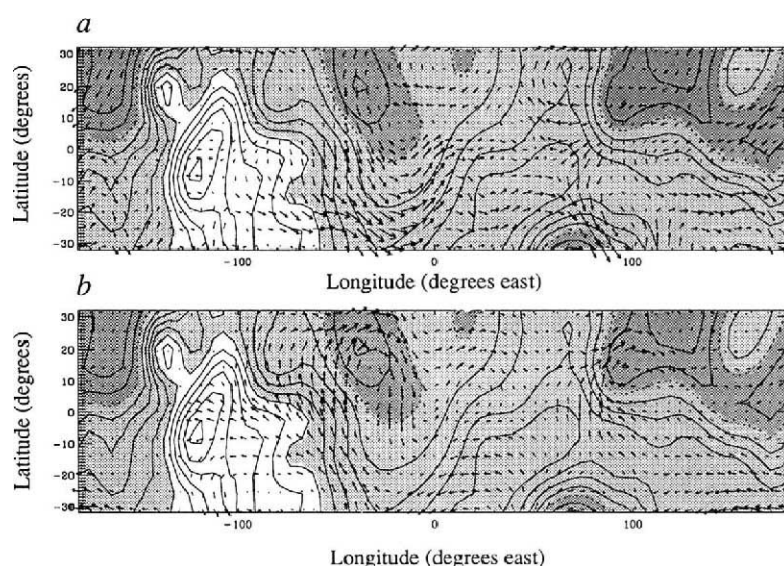
WESTERN boundary currents (WBCs) are an intensification of north–south flow adjacent to an eastward-facing meridional boundary. Although most familiar in the oceans (where the Gulf Stream is the best known example), WBCs also occur in the Earth's troposphere, the main example being the East African Jet<sup>1</sup>, which is thought to play an important role in the Asiatic monsoon. Here we identify boundary currents in a different geophysical context: a numerical simulation of the atmosphere of Mars. In our simulation, WBCs exist in association with significant cross-equatorial flow and the presence of equatorial martian topography, which has vertical scale far exceeding terrestrial relief<sup>2</sup>. The intensity and width of these currents depend on model parameters, notably the surface drag. From a comparison of our results with other martian models we suggest that WBCs have already been simulated, although they were not previously identified as such<sup>3</sup>. The available observational evidence appears to be consistent with the presence of martian WBCs, which may be important in the generation of global and great dust storms.

FIG. 1 *a, b*, Two longitude–latitude plots for Mars, showing 15 sol time-mean wind vectors (arrows) in the lowest model level superimposed on topography contours. The contours are at 1 km intervals, with the zero contour dotted. Regions below this are shaded dark grey, regions in the interval 0–5 km are shaded light grey and regions above 5 km elevation are unshaded. In both cases the longest arrow represents a velocity of  $30 \text{ m s}^{-1}$ . *a*, Results from a run simulating northern winter solstice, and *b*, results simulating southern winter solstice, both runs having been carried out with  $\tau = 10$  sol and T21 resolution which signifies model fields being truncated in spectral space so that the maximum retained total (longitudinal + latitudinal) wavenumber or spherical harmonic is 21. In both cases the larger of the two WBCs is at longitude  $50^\circ \text{ W}$  on the eastern flank of the Tharsis plateau, which stretches from  $50^\circ \text{ W}$  to  $120^\circ \text{ W}$ . Another smaller WBC is present on the eastern side of Syrtis Major, at  $80^\circ \text{ E}$ , this current being more intense in southern winter (*b*). A return flow is shown in the region  $0^\circ$ – $20^\circ \text{ W}$ . The major topographic features are Tharsis ( $130^\circ$ – $70^\circ \text{ W}$ ,  $20^\circ \text{ S}$ – $10^\circ \text{ N}$ ), Chryse Planitia (centred at  $40^\circ \text{ W}$ ,  $25^\circ \text{ N}$ ), Noachis (centred at  $10^\circ \text{ E}$ ,  $25^\circ \text{ S}$ ), Hellas ( $80^\circ \text{ E}$ ,  $30^\circ \text{ S}$ ) and Syrtis Major (centred at  $70^\circ \text{ E}$ ,  $10^\circ \text{ N}$ ).

The simplest theoretical context in which WBCs occur is represented by the linear frictionless shallow-water equations<sup>4</sup> in the presence of a meridional boundary, and in which the magnitude of the Coriolis parameter  $f$  varies with latitude  $y$  as  $f = f_0 + \beta y$ ,  $f_0$  and  $\beta$  being constants. On starting from rest and integrating forward in time, a steady state is reached everywhere except at the western boundary, where the width of the WBC produced will decay until infinitesimally small. In reality, of course, one of two things will happen to make the width of the jet tend to a non-zero limiting value: either friction dominates and the width of the jet tends approximately to  $O(1/\tau\beta)$ , where  $O$  means “of order” and  $\tau$  is the characteristic timescale over which the friction acts; or non-linear inertial effects dominate and the width of the jet tends to  $O((U/\beta)^{1/2})$  where  $U$  is a typical r.m.s. wind strength. Although such a model is much simpler than a real system, it can be helpful in determining whether a more realistic WBC is controlled by frictional or inertial forces. In the case of the equatorial region of Mars, where  $\beta \approx 4 \times 10^{-11} \text{ m}^{-1} \text{ s}^{-1}$  and  $1^\circ$  of longitude  $\approx 60 \text{ km}$ , it would be expected that a frictional WBC would have width  $O(5/\tau)^\circ$  of longitude, where  $\tau$  is measured in sols (1 sol is a martian day  $\approx 24.66 \text{ h}$ ) and an inertial WBC would have width  $O(3\sqrt{U})^\circ$  of longitude.

The numerical model used for this investigation is a simple general circulation model (SGCM) originally developed by Hoskins and Simmons<sup>5</sup>, which solves the hydrostatic primitive equations of dynamical meteorology on a sphere. In the vertical, the model uses the  $\sigma$ -coordinate system ( $\sigma$  is pressure divided by surface pressure) in finite-difference form<sup>6</sup> with 20 levels represented over the altitude range 0–80 km, whilst in the horizontal domain fields are represented by a truncated series of spherical harmonics<sup>5</sup>. Martian topography is derived from the US Geological Survey Digital Terrain Model (DTM)<sup>7</sup>. The SGCM represents atmospheric radiative processes by newtonian relaxation (on a timescale of 3 sol (ref.8)) to a specified zonal mean temperature state as commonly used in simple model studies of the Earth's atmosphere<sup>9</sup>. The temperature state used consists of a simple analytical function adapted to produce a time-mean state in the complete model which is similar to that of the NASA Ames GCM<sup>10</sup> and to observations<sup>11</sup> of the martian atmosphere from Mariner 9. Surface drag is represented by Rayleigh friction in the lowest model level on timescale  $\tau$ .

Simulations were carried out for atmospheric states representative of the two solstice seasons under low-dust conditions. We have studied the effect of varying surface drag over the range  $0.2 \text{ sol} < \tau < 10 \text{ sol}$  as well as the sensitivity to differences in both model and topographic resolution.



Figures 1 and 2 show the presence of equatorial WBCs in simulations with  $\tau = 10$  sol. One WBC lies on the eastern flank of the Tharsis Plateau, in Chryse Planitia at a longitude of  $\sim 50^\circ \text{ W}$ , as might be expected due to the large longitudinal gradient of topography there. The other lies on the eastern slopes of Syrtis Major, around longitude  $80^\circ \text{ E}$ , although its magnitude is smaller in the southern summer case. Both currents are thought to be WBCs because (1) they disappear in control simulations with no topography and (2) their width changes with surface drag in the manner expected for WBCs. As the strength

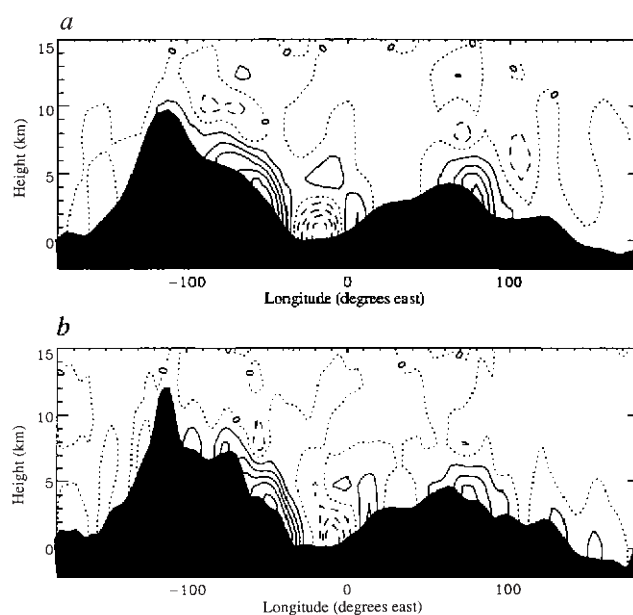


FIG. 2 A longitude–height plot for Mars, showing 15 sol time-mean meridional wind  $v$  (contours) in the lowest model  $\sigma$ -level ( $\sim 1 \text{ km}$  above the surface), along the equator, during southern winter solstice, showing the two boundary currents at  $50^\circ \text{ W}$  and  $80^\circ \text{ E}$ . *a* and *b* show results from runs carried out at T21 and T42 resolution respectively;  $\tau = 10$  sol in both cases. The return flow is evident at  $20^\circ \text{ W}$ . The difference between topography resolved at T21 and at T42 is at its greatest in the longitude range  $70^\circ$ – $120^\circ \text{ W}$  and within the altitude range 10–15 km, where  $v$  is low. The large vertical amplitude of the topography should again be noted. (Contour intervals,  $5.0 \text{ m s}^{-1}$ , the zero contour is dotted, negative contours are dashed.)



of the Hadley cell in simulations of northern and southern summers is about the same, WBC intensity is similar, and the highest winds occur in the summer tropics in both cases (Fig. 1). More sophisticated models show that the Hadley cell is stronger in southern summer<sup>10</sup>, which would be expected to lead to more intense WBCs during this season. During both solstices, the martian tropical zonal mean meridional circulation is dominated by a single cross-equatorial Hadley cell, as is the case at the Earth's solstices<sup>12</sup>. As the lower branch of this cell at the equator is concentrated into two WBCs, we expect martian WBCs to exert a strong influence on the flow within the Hadley cell.

Cross-equatorial flow can only happen when the flow departs from geostrophic balance (a simple balance between the pressure gradient and the Coriolis force on the flow), for example, owing to the effects of topography and/or surface friction. In our simulations, without topography, the meridional mass transport was augmented by  $\sim 8\%$  on increasing the drag from  $\tau = 10$  sol to  $\tau = 0.2$  sol because larger surface friction enhanced cross-equatorial flow. In the presence of topography, longitudinal pressure torques may enable net north-south flow, in this case in the form of WBCs. Meridional transport in the SGCM was found to increase by 11% when surface drag was reduced from  $\tau = 0.2$  sol to  $\tau = 10$  sol owing to the greater strength of WBCs in the latter case.

It was found that the nature of the current was highly dependent on the surface drag timescale used, firstly because reducing

surface drag caused higher surface winds and secondly because the nature of the WBC changed from frictional control at  $\tau = 0.2$  sol, to inertial control for  $\tau \geq 1$  sol (Fig. 3). The width of the jet reached a minimum value at  $\tau = O(1)$  sol (Fig. 3). For  $\tau = 10$  sol, a significant southward return flow also occurs at low levels near longitude  $20^\circ$  W, though this feature is not present for  $\tau = 2$  sol or less. The combined effect of narrowing and strengthening the WBC (each by a factor  $\sim 4$  in going from  $\tau = 0.2$  sol to  $\tau = 10$  sol, see Fig. 3) evidently results in a small net increase in the global meridional transport. As the martian Hadley cell is thought to be the most important mechanism for transporting quantities such as heat or trace constituents across the equator<sup>10</sup>, it must be concluded that WBCs play a large part in cross-equatorial transport of constituents, especially those, such as water vapour, that are expected to be concentrated near the surface<sup>13</sup>.

We noted in experiments at T42 resolution (where  $Tn$  denotes a spherical harmonic expansion truncated at a total, longitudinal and latitudinal, wavenumber  $n$ ) that the difference in circulation patterns and the strength of the Hadley cell between topography resolved at T21 and T42 respectively was negligible: WBCs

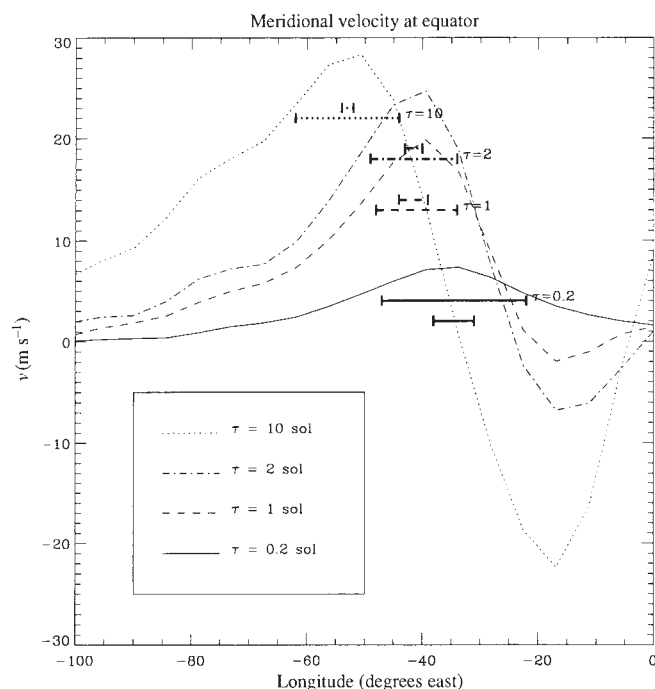


FIG. 3 The effect of surface drag on the simulation of the Tharsis WBC at southern winter solstice. All the simulations were carried out at model resolution T21, with values of  $\tau$  set at 0.2, 1, 2 and 10 sol. Each curve of meridional wind,  $v$ , is accompanied by two bars, the linestyle of the two bars being the same as the curve to which they apply (for example,  $\tau = 10$  is a dotted line and the two bars associated with it are dotted). Positive  $v$  indicates northward flow. The upper bar is the width the WBC would tend to if it was controlled purely by frictional forces and the lower bar is that expected if controlled purely by inertial forces. It can be seen that the width of the WBC produced with  $\tau = 0.2$  sol is consistent with friction control and the widths of the other three cases are consistent with inertial control. Note that the centre of the WBC seems to move westward as friction is reduced.

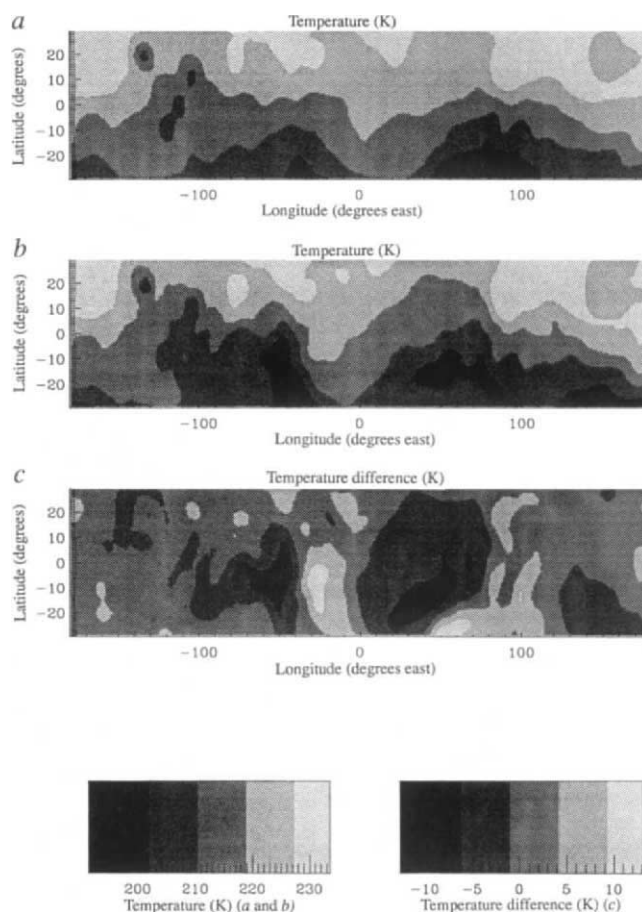


FIG. 4 Plots showing time-mean temperatures (shading) in the lowest model level. a, Run carried out at  $\tau = 0.2$  sol, and b, at  $\tau = 10$  sol. c, The difference  $b - a$ , caused mainly by advection of cold fluid northwards into the summer hemisphere by the Tharsis WBC at  $50^\circ$  W. (Note different temperature scale). The effect of the return flow is shown at  $25^\circ$  W and is evident as a hot area in c. The Syrtis Major WBC gives a far smaller signal. The model signal shown in the range  $60^\circ - 20^\circ$  W in c has a temperature of  $\sim 20$  K, reducing to  $\sim 2$  K at  $\sim 3.5$  km above the surface. This would be difficult to detect by conventional remote sounding instruments, unless it was possible to detect a corresponding time-mean surface temperature anomaly. It should certainly be measurable by a network of surface stations.

occurred in the same places and wind strengths were approximately the same. This is because the main differences in topography resolved at T21 and T42 occur in the altitude range 10–15 km, where meridional winds are low as it is the transition zone between the upper and lower branches of the Hadley cell (Fig. 2). Over a range of surface drag parameters the minimum width of WBCs were  $\sim 15^\circ$  in longitude. Although T21 resolution was adequate to resolve this, it was suspected that it may not have been enough to resolve the jet's core. We discovered, however, that the wind strengths were similar in runs at both T42 and T21 resolution (Fig. 2).

On Mars, bright depositional dust streaks observed at middle to low latitudes are believed to form in late southern summer and autumn<sup>14–16</sup>, and are thought to be representative of global scale winds<sup>3</sup> during these times. The orientation of these streaks, when averaged over  $5^\circ \times 5^\circ$  cells, imply north–south wind flow, and are especially coherent at  $40\text{--}50^\circ$  W (ref. 17), which is consistent with the longitude and direction of the Tharsis WBC produced in a model simulation during southern summer (Fig. 1a). In addition, Haberle *et al.*<sup>18</sup>, when comparing a one-dimensional boundary layer model with winds inferred from measurements taken during the entry of the Viking Lander 1 probe (at  $48^\circ$  W,  $22^\circ$  N) in early northern summer, found large discrepancies in wind speeds and their rotation with height which would be explained by significant advection of cold fluid from the southeast. This result is consistent with the presence of WBCs.

Boundary current winds also advect cold fluid towards the equator from the winter tropics, which might lead to a temperature signal observable by future spacecraft missions. The SGCM produces a strong thermal signal near the surface (see Fig. 4), although its magnitude may be significantly larger than that actually observed because the model representations of boundary-layer processes (such as heat and momentum transfer) are very simple at present.

WBCs may have a crucial role to play in the generation of global or great dust storms, because dust raising is critically dependent on low-altitude wind strength<sup>19</sup>, which in turn is the result of the superposition of winds caused by different forcing mechanisms<sup>20</sup>. This hypothesis is supported by the fact that large eastward-facing slopes in low latitudes are locations where global dust storms have commonly been seen to originate<sup>21</sup> and these locations are coincident with the WBCs in our model.

Finally, it should be noted that the large diurnal temperature variation on Mars<sup>22</sup> might cause a diurnal variation in the strength of the Hadley cell and hence in the strength of WBCs. Such effects are being investigated with a more sophisticated numerical model.  $\square$

Received 23 September 1993; accepted 4 January 1994.

1. Anderson, D. L. *T. Globl atmos. Res. Progr. Pubs Ser.* **23**, 317–355 (1980).
2. Esposito, P. *et al.* in *Mars* (eds Kieffer, H. H., Jakosky, B. M., Snyder, C. W. & Matthews, M. S.) Ch. 7 (Arizona Univ. Press, Tucson, 1992).
3. Greeley, R., Skyepeck, A. & Pollack, J. B. *J. geophys. Res.* **98**, 3183–3196 (1993).
4. Gill, A. E. *Atmosphere–Ocean Dynamics* 516–522 (Academic, San Diego, 1982).
5. Hoskins, B. J. & Simmons, A. J. *Quart. Jl. R. met. Soc.* **101**, 637–655 (1975).
6. Phillips, N. A. *J. Met.* **14**, 184–185 (1957).
7. *Atlas of Mars* (US geol. Surv. Topographic Ser. No. M 15M, Flagstaff, 1989).
8. Goody, R. & Belton, M. J. *Planet. Space Sci.* **15**, 247–256 (1967).
9. James, I. N. & Gray, L. J. *Q. Jl. R. met. Soc.* **112**, 1231–1250 (1986).
10. Haberle, R. M. *et al.* *J. geophys. Res.* **98**, 3093–3123 (1993).
11. Conrath, B. J. *et al.* *J. geophys. Res.* **78**, 4267–4278 (1973).
12. Lindzen, R. S. & Hou, A. Y. *J. atmos. Sci.* **45**, 2416–2427 (1988).
13. Jakosky, B. M. & Haberle, R. M. in *Mars* (eds Kieffer, H. H., Jakosky, B. M., Snyder, C. W. & Matthews, M. S.) Ch. 28 (Arizona Univ. Press, Tucson, 1992).
14. Greeley, R., Lancaster, N., Lee, S. & Thomas, P. in *Mars* (ed Kieffer, H. H., Jakosky, B. M., Snyder, C. W. & Matthews, M. S.) Ch. 22 (Arizona Univ. Press, Tucson, 1992).
15. Ward, A. W., Doyle, K. B., Helm, P. J., Weisman, M. K. & Witback, N. E. *J. geophys. Res.* **90**, 2038–2056 (1985).
16. Thomas, P., Ververka, J., Gineris, D. & Wang, L. *Icarus*, **60**, 161–179 (1984).
17. Kahn, R. A., Martin, T. Z., Zurek, R. W. & Lee, S. W. in *Mars* (eds Kieffer, H. H., Jakosky, B. M., Snyder, C. W. & Matthews, M. S.) 1042–1043 (Arizona Univ. Press, Tucson, 1992).
18. Haberle, R. M., Houben, H. C., Hertenstein, R. & Herdette, T. *J. atmos. Sci.* **50**, 1544–1559 (1993).
19. Greeley, R., Leach, R., White, B., Iversen, J. & Pollack, J. *Geophys. Res. Lett.* **7**, 121–124 (1980).
20. Leovy, C. B., Zurek, R. W. & Pollack, J. B. *J. atmos. Sci.* **30**, 749–762 (1973).
21. Martin, L. J. & Zurek, R. W. *J. geophys. Res.* **98**, 3221–3246 (1993).
22. Martin, T. Z. *Icarus* **45**, 427–446 (1981).

## Assessing the pyroclastic flow hazard at Vesuvius

Flavio Dobran\*†, Augusto Nerl††  
& Micol Todesco§

\* Department of Earth System Science, New York University,  
New York, New York 10003, USA

† Istituto Nazionale di Geofisica, Via di Vigna Murata 605, Rome, Italy

‡ GNV, § Dipartimento di Scienze della Terra, Università di Pisa,  
Via S. Maria 53, 56100 Pisa, Italy

**IN large eruptions, Vesuvius has generated catastrophic avalanches of tephra and hot gases, such as those that destroyed Pompei and Herculaneum in AD 79, and Torre del Greco and surrounding towns in 1631<sup>1–12</sup>. These avalanches (pyroclastic surges and flows) are produced from collapses of the eruptive column, and can travel at  $>100\text{ m s}^{-1}$ , with temperatures exceeding  $800^\circ\text{C}$ . In 1944 Vesuvius ended its most recent cycle of activity, which had begun with the explosive eruption of 1631. Here we use numerical simulations to assess the hazards posed by the pyroclastic flows that are likely to accompany the onset of the next cycle of activity. We examine three different scales of eruption, and use vent conditions established by modelling magma ascent along the conduit<sup>13,14</sup>. Our results indicate that large- and medium-scale eruptions can produce complete destruction in the 7 km radius around the volcano (an area in which one million people live and work) in about 15 minutes or less, and that only small-scale eruptions can be arrested by the topographic relief of Monte Somma.**

The Somma–Vesuvius complex has exhibited various types of activity over the past 35,000 yr (ref. 1). Large-scale plinian eruptions (Codola, Sarno, Basal, Greenish, Lagno Amendolare, Mercato, Avellino and Pompei) erupted several cubic kilometres of material and have occurred every few centuries to millennia, whereas medium-scale sub-plinian eruptions (AD 412 and 1631) have occurred every few centuries, with each erupting  $\sim 0.1\text{ km}^3$  of material<sup>15</sup>. It appears that the smaller-scale strombolian and effusive events occurring every few decades normally follow the plinian and sub-plinian eruptions until the conduit closes, and that the plinian and sub-plinian eruptions occur from the closed-conduit states of the volcano<sup>12</sup>. A common feature of the plinian eruptions is that they were interrupted intermittently as a result of partial column collapses producing pyroclastic surges and flows<sup>2,5</sup>, and ended with the interaction of magma with water from underground aquifers<sup>2,9,16</sup>. The plinian and sub-plinian eruptions from the Avellino eruption  $\sim 3,400$  yr ago to AD 1631 are all characterized by the emission of highly differentiated magmas, trachytic or phonolitic in character<sup>17,18</sup>. The pumice-fall deposits of Avellino, Pompei and the AD 1631 eruptions consist of white phonolite at the base and grey tephritic phonolite at the top<sup>11,19</sup>. Some insight on the future activity of Vesuvius, and on its effects on the surrounding area, can be established from studies of its past behaviour and distribution of its products<sup>19</sup>. The results from this kind of study are, however, poorly constrained by thermodynamics, geophysics and thermofluid-dynamics, and a more precise assessment of the future behaviour of Vesuvius could be achieved by modelling the entire volcanic system<sup>12</sup>.

The available volcanological and petrological data for Vesuvius have recently been used to model magma ascent during the AD 79 eruption<sup>13,20</sup>. Based on the magma composition, temperature and crystal content, location of magma chamber, and stratigraphy, this modelling yielded likely values of conduit diameter, gas–pyroclast pressure and velocity distributions as a function of the mass discharge rate. The gas–pyroclast conditions at the vent were subsequently used in a physical model of the volcanic column<sup>14</sup> to study pyroclastic flows caused by column collapses. This column model employs a subgrid-scale



turbulence model, and the thermo-fluid dynamics includes the dispersive pressure produced by solid-particle collisions modelled by kinetic theory. The flow above the vent is assumed to be axisymmetric, consisting of a single particle size, and involving water vapour that may mix with the temperature- and pressure-stratified atmosphere.

The magma ascent and volcanic column physical models<sup>13,14</sup>, combined with different topographic profiles of Vesuvius, provide a state-of-the-art simulation of the temporal and spatial evolution of eruption columns and associated pyroclastic flows. The main limitation of this axisymmetric flow modelling, as well as of that used in previous volcanic column studies<sup>21,22</sup>, is that it does not account for the presence of valleys along the slopes of the volcano where three-dimensional topographic effects may strongly control the propagation of pyroclastic flows. Field studies of volcanic deposits and direct observations, however, distinguish dense and less mobile pyroclastic flows from less dense and more mobile pyroclastic surges<sup>23</sup>. This distinction is usually based on textural features of pyroclastic deposits which the model cannot resolve because of the use of a single particle size, large computational grid sizes and short simulation times. The characteristics of the simulated gas-pyroclast dispersions suggest, however, that they may be associated more closely with the pyroclastic surges. Because of the absence of precise quantitative criteria for distinguishing flows from surges, these flows will thus be referred to as 'pyroclastic flows'.

The input data for magma ascent and volcanic column models are shown in Table 1 as a function of three different mass eruption rates representing three different scales of eruption. The large mass eruption rate represents the grey eruption phase of Vesuvius in AD 79 (ref. 10), whereas the medium mass eruption rate is typical of a sub-plinian event such as that of AD 1631. The small mass eruption rate is an order of magnitude smaller than the large discharge rate. The gas-pyroclast conditions at the vents reported in the table pertain to the AD 79 grey magma composition and crystal content<sup>13,16,24</sup>, and to a dissolved water content of 2 wt%. This magma water content reflects the more recent data of 2.5 wt% (ref. 25) pertaining to the AD 79 grey eruption phase, but some simulations were also performed with a water content of 3.5 wt% as reported previously in the literature<sup>10</sup>. Except as described below, detailed variations in magma composition (which at Vesuvius are associated with different scales of eruptions) are not taken into account in the model to limit the number of parameters. The conduit lengths and associated magma chamber pressures in Table 1 reflect

TABLE 1 Input data for magma ascent and volcanic column modelling

| Scale of the eruption                                | Large             | Medium            | Small             |
|------------------------------------------------------|-------------------|-------------------|-------------------|
| Mass eruption rate ( $\text{kg s}^{-1}$ )            | $1.5 \times 10^8$ | $5.0 \times 10^7$ | $1.5 \times 10^7$ |
| Conduit length (km)                                  | 5                 | 3                 | 3                 |
| Magma chamber pressure (MPa)                         | 124.7             | 74.7              | 74.7              |
| Magma chamber temperature (K)                        | 1,123             | 1,123             | 1,123             |
| Pyroclast size ( $\mu\text{m}$ )                     | 100               | 100               | 100               |
| Dissolved water content of magma (wt%)               | 2                 | 2                 | 2                 |
| Conduit diameter (m)                                 | 100               | 60                | 40                |
| Gas velocity at the vent ( $\text{m s}^{-1}$ )       | 118               | 130               | 135               |
| Pyroclast velocity at the vent ( $\text{m s}^{-1}$ ) | 110               | 115               | 116               |
| Gas-pyroclast temperature at the vent (K)            | 1,123             | 1,123             | 1,123             |
| Pressure at the vent (MPa)                           | 1.3               | 0.96              | 0.65              |
| Particle volumetric fraction at the vent             | 0.067             | 0.052             | 0.0365            |

The mass eruption rate conduit length, magma chamber pressure and temperature, pyroclast size after magma fragmentation, and the AD 79 grey magma composition, crystal and dissolved water content, form the input data for magma ascent modelling<sup>13</sup>. The conduit diameter, pyroclast size and volumetric fraction, and gas-pyroclast temperature, pressure and velocities at the vent were established from magma ascent modelling and form the input data for the volcanic column modelling.

different locations of plinian (large) and sub-plinian (medium and small) magma chambers<sup>8,12</sup>, whereas the conduit diameters were determined by assuming that the given mass eruption rates correspond to choked or maximum flow conditions at the exit of the vent<sup>20</sup>. The pyroclast size of 100  $\mu\text{m}$  employed in computer simulations represents one of the more mobile parts of the flow, as larger pyroclasts tend to be deposited at shorter distances from the vent whereas the smaller ones tend to follow the ascending hot plumes forming above the collapsed portion of the column and above the pyroclastic flow<sup>14,26,27</sup>.

The vent conditions in Table 1 produced collapsed volcanic columns which generate pyroclastic flows along the two investigated topographic profiles of Vesuvius (Figs 1 and 2). Topography A accounts for the southern slope of the volcano toward the Tyrrhenian Sea, whereas topography B represents the route

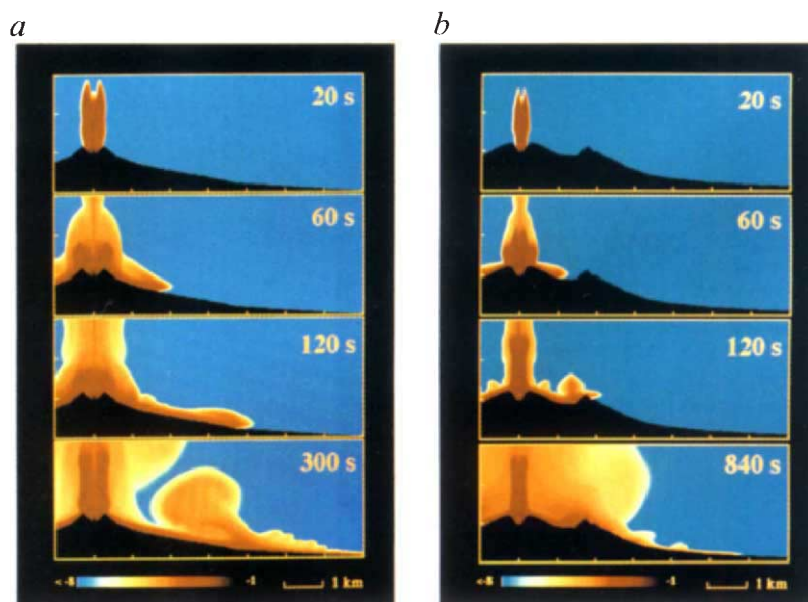


FIG. 1 a, Distribution of particle volumetric fraction at 20, 60, 120 and 300 s along the slope of Vesuvius toward the Tyrrhenian Sea (profile A) after the initiation of gas-particle discharge from a large-scale volcanic eruption of Vesuvius. b, Distribution of particle volumetric fraction at 20, 60, 120 and 840 s toward Somma Vesuviana (profile B) after the initiation of gas-particle discharge from a medium-scale volcanic eruption of Vesuvius. The red colour denotes very high, and the blue colour very low, particle volumetric fractions. The numbers on the colour scale at the bottom of the figures represent exponents to the base 10.

over Monte Somma towards the north in the direction of Somma Vesuviana. Figure 1a illustrates the evolution in time of the distribution of particle volumetric fraction of the large-scale eruption along the topography A. At  $\sim 20$  s after the beginning of the eruption, the volcanic column reaches a height of  $\sim 3$  km above the vent. Thereafter it begins to spread radially due to the loss of vertical momentum, and collapses to form a pyroclastic flow while a low-density plume rises above the collapsed portion of the column. At 60 s, the ground-hugging flow reaches a distance of  $\sim 2$  km from the vent. At 120 s, the flow reaches a distance of 4 km, and at 300 s, it enters the Tyrrhenian Sea 7 km away from the vent. The results from computer simulations employing topography B (Fig. 1b) show that the morphological barrier of Monte Somma cannot stop the propagation of such a pyroclastic flow on the northern slope. By 300 s, a large-scale volcanic event in the vesuvian area is thus capable of destroying completely all the surrounding towns in the vicinity. The effect of Monte Somma on the pyroclastic flow propagation of a medium-scale eruption is shown in Fig. 1b by a sequence of time frames, demonstrating that even in this case the flow is not arrested and that it reaches a distance of  $\sim 6$  km in 1,000 s. As seen in Fig. 1a, the break-in-slope of the volcano at  $\sim 3.5$  km from the vent in the direction of the Tyrrhenian Sea produces a rising cloud (phoenix or co-ignimbrite column<sup>14</sup>) above the flow which subsequently merges with the collapsing and rising portions of the main column, and begins forming a large inward-necking volcanic plume. The Monte Somma relief shown in Fig. 1b produces an earlier development of the phoenix column (at  $\sim 1.2$  km from the vent) beyond which a small pyroclastic flow propagates along the northern slope. Computer simulations of pyroclastic flows were not carried out beyond the 7-km radius from the vent in order to keep the calculations at a manageable level.

On the basis of these results, an estimate of the pyroclastic flow hazard at Vesuvius can be made in terms of the different scales of the eruption, topographies A and B, and distances reached by the flow at different times (Fig. 2). As indicated above, the flow from a large-scale eruption cannot be stopped by any morphological barrier surrounding the volcano, and the effect of Monte Somma is only to retard the flow arrival times at sites distant from the vent (Fig. 2a). For a medium-scale volcanic event this flow retardation is more pronounced and is approximately twice as large as that of a large-scale eruption (Fig. 2b). The pyroclastic flow from a small-scale eruption also

#### BOX 1 The AD 79 and 1631 eruptions of Vesuvius

The AD 79 Pompei eruption of Vesuvius started at about 13:00 on 24 August with the formation of a plinian column that was preceded by a phreatomagmatic opening phase. For the following seven hours, a phonolitic magma (white pumice) was ejected. At about 20:00 on the same day, the magma composition changed to tephritic phonolite (grey pumice), and after 01:00 on 25 August several pyroclastic surges and flows occurred, some of which were directly responsible for the destruction of Herculaneum and Pompei where 2,000 people were killed<sup>1,3,5,7,10</sup>. The Pompei eruption discharged  $\sim 4$  km<sup>3</sup> of dense rock equivalent<sup>5</sup>. The AD 1631 eruption of Vesuvius was the most destructive event in the recent history of this volcano<sup>11</sup>. On 16 December 1631 at 06:30, the eruption started with the ejection of gas and ash, and by 10:00 it produced a plinian column. Between 19:00 and 22:00 on the same day, the eruption produced a succession of seismic shocks, and by 02:00 on 17 December a glowing cloud was seen issuing from the summit crater, flowing into the valley between the cone of Vesuvius and Monte Somma. During the night, lahars were also seen coming down from the northern slopes of Monte Somma, devastating the land around Ottaviano. At 11:00 a violent earthquake occurred and the central crater was seen discharging ash, gas and stones, which poured down the slopes of the cone of Vesuvius. These pyroclastic flows descended from the mountain in lobes and destroyed Torre del Greco and Torre Annunziata, 6 and 7 km from the crater, respectively. The eruption produced decapitation of the cone by  $\sim 500$  m and  $\sim 4,000$  deaths, primarily because of the pyroclastic flows. Extensive destruction was suffered by Ottaviano, Massa di Somma, San Sebastiano, San Giorgio and other places, and  $\sim 500$  km<sup>2</sup> were covered by ash which destroyed crops, vineyards and cattle. □

reaches the Tyrrhenian Sea at  $\sim 1,000$  s but is arrested by Monte Somma (Fig. 2c). The computer simulations for a topography toward Ottaviano also demonstrate that pyroclastic flows from small-scale eruptions are arrested by Monte Somma. It is necessary to stress, however, that on the southern slopes of the volcano even small-scale eruptions can produce pyroclastic flows which reach the Tyrrhenian Sea, and which would produce localized destruction and panic among the population. When the computer simulations were performed with 3.5 wt% of dissolved water in the magma and for topography A, it was also found that the columns collapsed, but that the flows required slightly

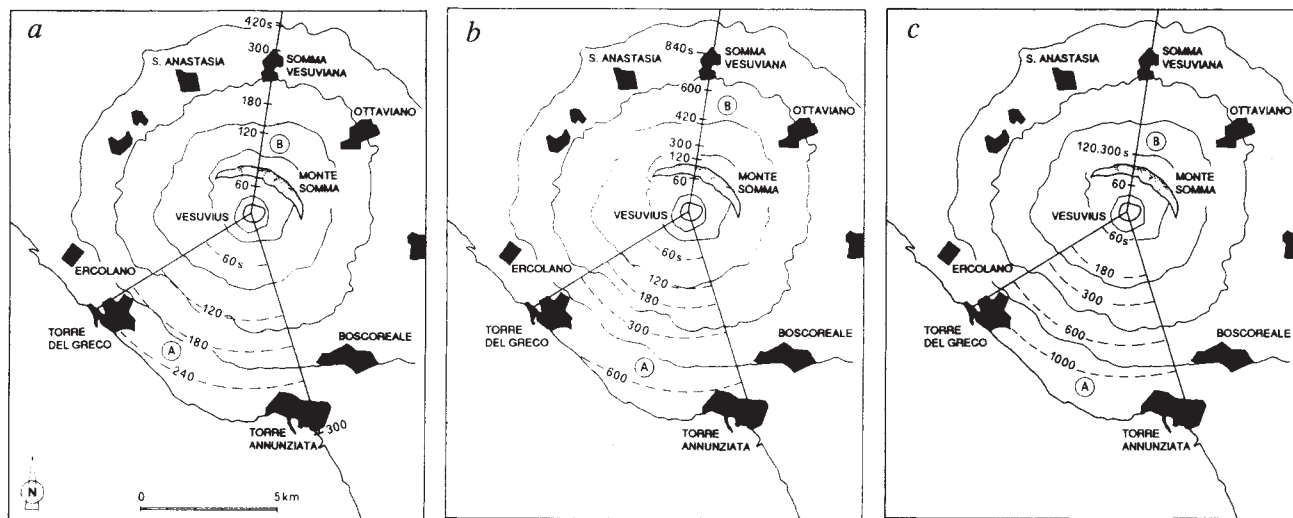


FIG. 2 Arrival time(s) of pyroclastic flows at different distances from the vent, for large- (a), medium- (b), and small-scale (c) eruptions, and topographies A and B in the directions of Tyrrhenian Sea and Somma Vesuviana, respectively. Note that the pyroclastic flows from all erup-

tions reach the Tyrrhenian Sea, and that only the flow from the small-scale eruption is arrested by Monte Somma. The hazard can be assessed by combining this information with the amount of material deposited, which can be derived from Fig. 1.



longer times to reach the Tyrrhenian Sea than the flows produced with 2 wt% of dissolved water in magma.

The results from computer simulations are consistent with the volcanological evidence of the pyroclastic flows and surges during the AD 79 and 1631 eruptions of Vesuvius<sup>5,10,11</sup>. In AD 79, the pyroclastic flows and surges reached the Tyrrhenian Sea in few minutes; a surge reached Somma Vesuviana to the north and Ottaviano to the northeast of Monte Somma. These pyroclastic flows and surges are generally associated with the partial collapses of the column produced by the changing conditions at the vent during the withdrawal of grey magma from magma chamber. According to a previous interpretation of the AD 79 eruption at Vesuvius<sup>5</sup>, the main thrusting part of the column (the 'jet') extended ~3 km above the crater, only the jet part of the column collapsed during pyroclastic flow and surge productions, and the time interval between generations of some surges lasted only few minutes. These considerations are also consistent with our computer simulations. The pyroclastic flows and surges during the sub-plinian eruption of Vesuvius in AD 1631 also reached the Tyrrhenian Sea as predicted by our computer simulations for a medium-scale eruption. During the AD 79 eruption, pyroclastic flows and surges were produced only at later stages of the grey magma eruption phase<sup>5,7</sup>. To verify the effects of changing magma composition on the column collapse, we simulated the magma ascent and volcanic column for the white magma composition and 4.7 wt% of dissolved water in magma<sup>5</sup>, followed by the emission from the same vent diameter of grey magma composition containing 3.5 wt% of dissolved water. This initially produced a plinian column, but with the emission of the grey magma the column exhibited partial collapses, implying that our simulations of pyroclastic flows at Vesuvius reflect the later stage of the grey magma eruption phase.

It should be noted that the predicted pyroclastic flow and surge distributions in the vesuvian area assume axisymmetric flow conditions and one granulometric particle size class. This implies that in reality these simulations should underestimate the maximum potential hazard due to these flows. The real flows are directed along preferential directions determined by local topography, and different sizes of pyroclasts will produce different mobilities, implying that surges, but not flows, may cross high reliefs as happened in the AD 79 eruption. Moreover, the presence of prevailing winds during eruptions can affect the column stability and also direct pyroclastic flows along preferred directions. These computer simulations of pyroclastic flows for the first time illustrate the timescales associated with this type of volcanic hazard in the vesuvian area. These simulations also indicate that future explosive eruptions at Vesuvius will have catastrophic effects for several hundred thousand people who live along its slopes, unless effective evacuation plans and new roads are built, and people begin depopulating the area. A proper way to assess the future volcanic hazard in this area is to develop a Global Volcanic Simulator for Vesuvius<sup>12,28</sup>, based on interdisciplinary data and modelling all relevant physical processes in the volcanic complex. The simulator could be used to reconstruct more accurately past volcanic events at Vesuvius, forecast future eruptions, establish new safety codes for building and road construction, determine the best evacuation routes and educate the local population by dramatizing the impending danger of the sleeping giant. □

Received 24 August; accepted 3 December 1993.

1. Lirer, L., Pescatore, T., Booth, B. & Walker, G. P. L. *Geol. Soc. Am. Bull.* **84**, 759–772 (1973).
2. Sheridan, M. F., Barberi, F., Rosi, M. & Santacroce, R. *Nature* **289**, 282–285 (1981).
3. Sigurdsson, H., Cashdollar, S. & Sparks, R. S. J. *Am. J. Archeol.* **86**, 39–51 (1982).
4. Santacroce, R. *J. Volcan. geotherm. Res.* **17**, 237–248 (1983).
5. Sigurdsson, H., Carey, S., Cornell, W. & Pescatore, T. *Nat. Geogr. Res.* **1**, 332–387 (1985).
6. Arnó, V. et al. in *Somma-Vesuvius* (ed. Santacroce, R.) Vol. 114, 53–103 (CNR Quaderni, Roma, 1987).
7. Carey, S. & Sigurdsson, H. *Geol. Soc. Am. Bull.* **99**, 303–314 (1987).
8. Santacroce, R. (ed.) *Somma-Vesuvius* Vol. 114 (CNR Quaderni, Roma, 1987).
9. Barberi, F. et al. *J. Volcan. geotherm. Res.* **38**, 287–307 (1989).
10. Sigurdsson, H., Cornell, W. & Carey, S. *Nature* **345**, 519–521 (1990).

11. Rosi, M., Principe, C. & Vecchi, R. *J. Volcan. geotherm. Res.* (in the press).
12. Dobran, F. *Global Volcanic Simulation of Vesuvius* (Giardini, Pisa, 1993).
13. Papale, P. & Dobran, F. *J. Volcan. geotherm. Res.* **58**, 101–132 (1993).
14. Dobran, F., Neri, A. & Macedonio, G. *J. geophys. Res.* **98**, 4231–4259 (1993).
15. Macedonio, G., Pareschi, M. T. & Santacroce, R. *J. Volcan. geotherm. Res.* **40**, 327–342 (1990).
16. Barberi, F., Navarro, J. M., Rosi, M., Santacroce, R. & Sbrana, A. *Rend. Soc. Ital. Miner. Petrol.* **43**, 901–926 (1988).
17. Civetta, L., Galati, R. & Santacroce, R. *Bull. volcan.* **53**, 287–300 (1991).
18. Civetta, L. & Santacroce, R. *Acta vulcan.* **2**, 147–159 (1992).
19. Rosi, M., Santacroce, R. & Sheridan, M. in *Somma-Vesuvius* (ed. Santacroce, R.) Vol. 114, 197–220 (CNR Quaderni, Roma, 1987).
20. Dobran, F. *J. volcan. geotherm. Res.* **49**, 285–311 (1992).
21. Wohletz, K. H., McGetchin, T. R., Sandford, M. T. & Jones, E. M. *J. geophys. Res.* **89**, 8269–8285 (1984).
22. Valentine, G. A., Wohletz, K. H. & Kieffer, S. W. *Geol. Soc. Am. Bull.* **104**, 154–165 (1992).
23. Cas, R. A. F. & Wright, J. V. *Volcanic Successions: Modern and Ancient* (Chapman & Hall, London, 1987).
24. Barberi, F. et al. *Bull. Volcan.* **44**, 295–315 (1981).
25. Cioni, R. et al. (abstr.) Annual Meeting CNR-Gruppo Nazionale per la Vulcanologia, 8–10 June, Rome (Giardini, Pisa, 1993).
26. Neri, A. & Dobran, F. *J. geophys. Res.* (submitted).
27. Giordano, G. & Dobran, F. *J. Volcan. geotherm. Res.* (in the press).
28. Dobran, F. *Int. Symp. on Large Explosive Eruptions* (Accademia Nazionale dei Lincei, 24–25 May, Rome (1993).

## Parasitism, mutation accumulation and the maintenance of sex

R. Stephen Howard & Curtis M. Lively

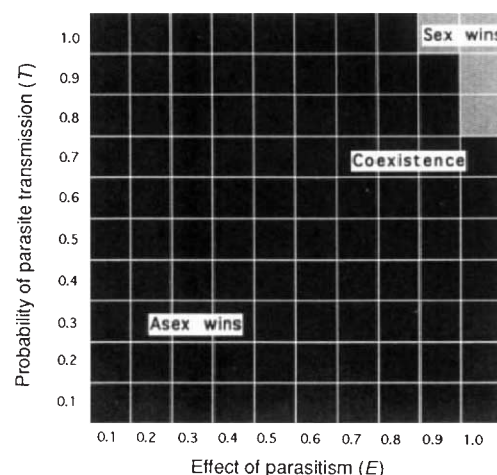
Department of Biology, Indiana University, Bloomington, Indiana 47405, USA

Two classes of models attempt to explain why obligate parthenogenesis only rarely replaces sexual reproduction in natural populations, in spite of the apparent reproductive advantage that parthenogens gain by producing only female offspring<sup>1</sup>. The mutation-accumulation models suggest that sex is adaptive because it purges the genome of harmful recurrent mutations<sup>2,3</sup>. The ecological genetic models postulate that sex is adaptive in variable environments, particularly when the relevant variation is generated by coevolutionary interactions with parasites<sup>4–7</sup>. Both of these models have considerable merit, but would seem to have limitations. The mutation-accumulation models require high rates of mutation<sup>3,8</sup>; the coevolutionary models require that parasites have severe fitness effects on their hosts<sup>9</sup>. In addition, parasites could select for clonal diversity and thereby erode any advantage that sex gains by producing variable progeny<sup>10</sup>. Here we consider the interaction between mutation accumulation and host-parasite coevolution. The results suggest that even moderate effects by parasites combined with reasonable rates of mutation could render sex evolutionarily stable against repeated invasion by clones.

The crux of the 'red queen' hypothesis is that there is time-lagged, frequency-dependent selection against common host genotypes<sup>5–7</sup>. Parasites are expected to evolve rapidly so as to infect disproportionately any genotype that becomes common, and thereby drive it down in frequency<sup>11,12</sup>. This sets up an oscillation in which parasite gene frequencies track the gene frequencies of their host. This kind of antagonistic coevolution could theoretically prevent the local fixation of any host genotype; but, to prevent the fixation of a clone with a twofold reproductive advantage, the direct effects of parasites would need to be severe<sup>9</sup>, or there would have to be rank-order truncation selection against the most infected individuals<sup>13</sup>.

Even with severe effects of parasites (for example, parasitic castration), it would seem that clones would still have a reproductive advantage when rare. Such an advantage could easily explain the local coexistence of reproductively isolated sexual and parthenogenetic lines, but it leads to another potential problem. Repeated mutations to parthenogenetic reproduction would lead to the accumulation of clones; and assuming that

FIG. 1 Results from computer simulations in which sexual populations were challenged by asexual lineages in the presence of coevolving parasites. Each block in the grid represents the majority outcome from five replicate runs of the simulation for a single combination of parasite transmission probability ( $T$ ) and effect of parasitism ( $E$ ). Coexistence was defined as the persistence of both types for more than 300 generations. At the beginning of each simulation, 500 unlinked loci in each of 1,000 sexual individuals received a mutation with probability  $U/s/500$ . The fitness of an individual with  $k$  mutations was calculated as  $(1-s)^k$ . Thus, the sexual population was initialized with the equilibrium mean ( $U/s$ ) and variance for mutation number. A clonal genotype of 20 individuals was similarly initialized with  $U/s$  mutations. After initialization, both sexual and asexual hosts randomly accumulated mutations each generation with a Poisson mean of  $U$ . There were two parasite generations for each host generation. The genetic interaction was simulated by considering two unlinked loci, each with two alleles, for both host and parasite. To maintain variation in the parasite population, mutation between allelic forms occurred with a probability of 0.03. For each parasite generation, hosts were drawn individually and exposed to a randomly selected parasite genotype with probability  $T$ . If exact matching occurred at both loci, the host was marked as infected and the parasite entered into the pool of reproductives. If there was less than complete matching, the parasite died. Allele frequencies for host and parasites were all initialized at 0.5. Host reproduction was simulated by drawing individuals at random with replacement. If the selected individual was sexual, a second individual was randomly selected for cross-fertilization. Each individual then produced 5 embryos if they were uninfected and an average of  $5(1-E)$  embryos if infected. The haploid embryos were then randomly assigned, at each locus, the



allele from one of the two parents. If, instead, the selected individual was parthenogenetic, 20 embryos identical to uninfected parents were produced; and, on average,  $20(1-E)$  embryos identical to infected parents were produced. Reproduction was simulated in this way until the number of broods produced equalled the total number of adults in the population. A maximum of 1,000 juveniles were then selected randomly to become the next generation of adults.

these clones are random samples of the local genetic variation for parasite recognition, multiple clonal lineages with different recognition phenotypes would be expected to result. Thus clones should accumulate, and the advantage of cross-fertilization would be expected to diminish as the number of clones increases. If this reasoning is correct, there must be some mechanism, in addition to parasites, that removes clones at least as fast as they are generated. One such mechanism could be mutation accumulation in the clonal lineages.

We have evaluated the interacting effects of parasites and the accumulation of mutations through the 'ratchet'-like process

first noted by Muller<sup>2</sup>, and later formalized by Felsenstein<sup>14</sup> and Haigh<sup>15</sup>. Muller's idea was that in finite populations there would be a good chance for the stochastic loss of the clonal class that is the least loaded with mutations. Computer simulations showed that, in the absence of mutation, parasites generated a decisive advantage to sex only when parasite transmission probabilities were high ( $>70\%$ ) and the effects of parasites on host fitness were dire ( $>80\%$  loss of fitness), which accounts for only 4% of the possible parameter space (Fig. 1). This result is consistent with May and Anderson's<sup>9</sup> contention that antagonistic coevolution favours sex only when parasites have severe effects on host

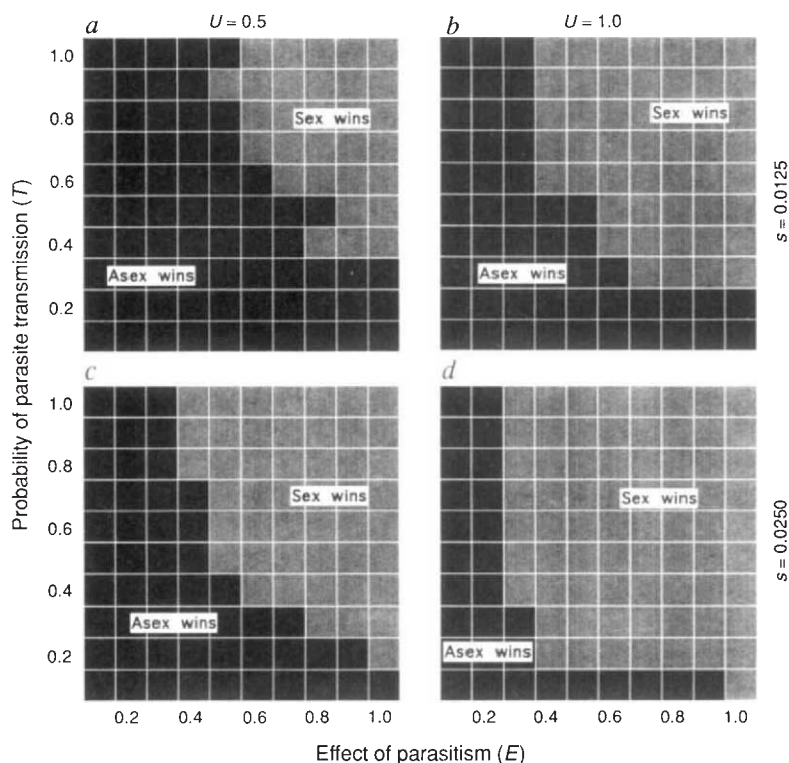


FIG. 2 Results from computer simulations in which sexual populations were challenged by monotypic asexual lineages in the presence of coevolving parasites and selection against harmful recurrent mutations. Each block in the grids represents the majority outcome from five replicate runs of the simulation for a single combination of probability of parasite transmission ( $T$ ) and the effect of parasitism ( $E$ ). The simulations were run for four combinations of mutation rate ( $U$ ) and effect of mutation ( $s$ ): a,  $U = 0.5$  and  $s = 0.0125$ ; b,  $U = 1.0$  and  $s = 0.0125$ ; c,  $U = 0.5$  and  $s = 0.025$ ; d,  $U = 1.0$  and  $s = 0.025$ . Note that the range of parameter space for which sex is evolutionarily stable to invasion by parthenogenesis increases as the mutation rate increases, and as selection against mutations increases.



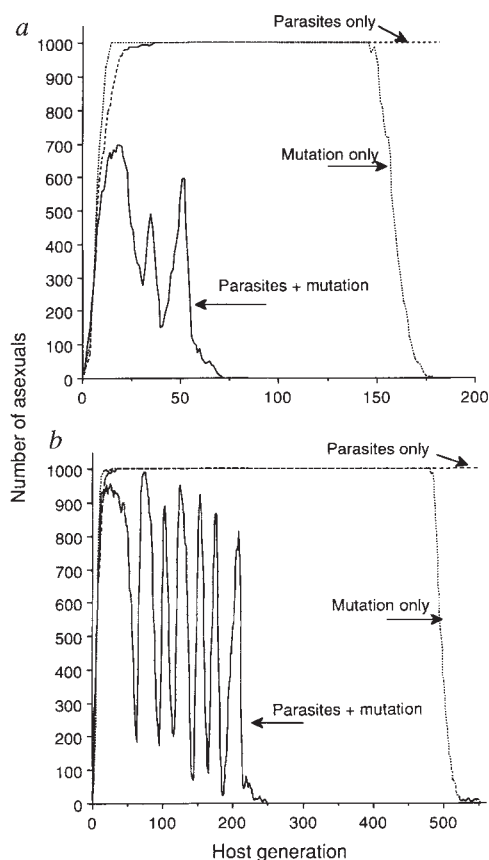


FIG. 3 Population dynamics for monotypic asexual lineages in competition with sexual populations in the presence of either parasites or mutation, or both. **a**, Parameters for these runs included a mutation rate ( $U$ ) of 0.5 per genome per generation, selection coefficient ( $s$ ) = 0.025, probability of parasite transmission ( $T$ ) = 0.5, and effect of parasites on host fitness ( $E$ ) = 0.5. Note that, for mutation only, the clone fixes in a population of 1,000 individuals in less than 50 generations; the clone then undergoes 'mutational meltdown' (in the sense of Lynch and Gabriel<sup>16</sup>), beginning at about generation 480. For parasites only, the clone also fixes in about 50 generations, but it does not later undergo the meltdown. Finally, for mutations combined with parasites, the clonal population does not replace the sexual population, but rather it oscillates. Each oscillation increases the mutational load through the action of Muller's ratchet, and there is a tendency for the lowest point of the oscillation to decline with time. Finally, the clone begins the mutational meltdown (see ref. 16) at about generation 220, and it is quickly eliminated from the population. **b**, As for **a**, except that the mutation rate  $U$  has been increased from 0.5 to 1.0. Note the more rapid extinction of the asexual lineage as compared with **a**.

fitness. There was, however, a fairly large region of parameter space (25%) defined by more intermediate values under which sexual and parthenogenetic populations coexisted, but this coexistence would be expected to be temporary, as selection for clonal diversity in this region would eventually erode the advantages that sex gains through the production of variable offspring.

The region of persistence by the sexual population increased dramatically by the addition of mutation to the model (Fig. 2). For example, for moderate mutation rates per genome ( $U = 0.5$ ;  $U$  is the Poisson-distributed mutation rate per haploid genome per generation) and weak selection against mutation ( $s = 0.0125$ ;  $s$  is the effect of a single mutation), sex was stable over 29% of the total parameter space. The region for coexistence of sexual and parthenogenetic populations for these values shrank to 9% of the total parameter space (Fig. 2a). Holding selection constant, but increasing the mutation rate to  $U = 1.0$ , sex was stable over 49% of the total parameter space, and the region of coexistence of sexual and asexual populations was completely eliminated (Fig. 2b). An almost identical result was gained by holding the per genome mutation rate at 0.5, but increasing the selection against mutations to  $s = 0.025$ ; sex was stable over 48% of the parameter space, and there was no region of stable coexistence (Fig. 2c). Finally, for mutation rates of  $U = 1.0$  and selection coefficients of  $s = 0.025$ , sex was stable over 71% of the total parameter space (Fig. 2d). Hence, host-parasite coevolution could interact with Muller's ratchet to confer evolutionary stability to sexual reproduction.

Figure 3 shows the effect of the interaction between parasites and mutation. At generation zero, a rare mutant that reproduces by parthenogenesis is introduced into a sexual population ( $N = 1,000$ ). The mutant spreads rapidly, but is knocked back at about generation 20 by time-lagged frequency-dependent selec-

tion from the parasite. As expected, the mutant clone then oscillates in density over time. However, each time the clone is depressed in number by the parasite, it becomes more susceptible to the accumulation of mutations by Muller's ratchet. (We use Muller's ratchet here to include both the gain of mutations through the stochastic loss of the least-loaded subclones and the probabilistic gain of mutations through mutation pressure.) Note that in Fig. 3, because of the accumulation of mutations, the trough of each oscillation by the clone becomes lower with time. Eventually the clone cannot replace itself and 'melts down'<sup>16</sup>. Hence, there is short-term coexistence of the sexual and parthenogenetic host populations due to the effects of the parasite, but mutation accumulation eventually eliminates the clone. Sex would be expected to be stable in the long term if clones are eliminated by this process at least as fast as they are generated by mutation.

Thus, moderate effects by parasites combined with reasonable rates of mutation could render sex evolutionarily stable against repeated invasion by clones. Parasites prevent the fixation of clones and the elimination of sex in the short term, and they increase the rate at which clones accumulate mutations through the action of Muller's ratchet by periodically depressing the population size of the clone. Mutation accumulation eventually leads to the extinction of clones. These results suggest that, where sexual and asexual populations coexist, the individual clones are transitory. □

Received 9 September; accepted 19 November 1993.

1. Maynard Smith, J. *The Evolution of Sex* (Cambridge Univ. Press, UK, 1978).
2. Muller, H. J. *Mutat. Res.* **1**, 2-9 (1964).
3. Kondrashov, A. S. *Nature* **336**, 435-440 (1988).
4. Levin, D. Am. Nat. **109**, 437-451 (1975).

5. Jaenike, J. *Evol. Theory* **3**, 191–194 (1978).
6. Hamilton, W. D. *Oikos* **35**, 282–290 (1980).
7. Bell, G. *The Masterpiece of Nature* (Univ. California Press, Berkeley, 1982).
8. Howard, R. S. *Theor. Populat. Biol.* (in the press).
9. May, R. M. & Anderson, R. M. *Proc. R. Soc. B* **219**, 281–313 (1983).
10. Lively, C. M. *Evolution* **46**, 907–913 (1992).
11. Clarke, B. in *Genetic Aspects of Host-Parasite Relationships* vol. 14 (eds Taylor, A. E. R. & Muller, R.) 87–103 (Blackwell Scientific, Oxford, 1976).
12. Hutson, V. & Law, R. *Proc. R. Soc. B* **213**, 345–359 (1981).
13. Hamilton, W. D., Axelrod, R. & Tanese, R. *Proc. natn. Acad. Sci. U.S.A.* **87**, 3566–3573 (1990).
14. Feisenstein, J. *Genetics* **78**, 737–756 (1974).
15. Haigh, J. *Theor. Populat. Biol.* **14**, 251–267 (1978).
16. Lynch, M. & Gabriel, W. *Evolution* **44**, 1725–1737 (1990).

ACKNOWLEDGEMENTS. We thank L. Delph, G. Hudock, S. Johnson and O. Sayeed for helpful suggestions, and the Centre for the Integrative Study of Animal Behaviour for computer resources. This study was supported by the US NSF.

## Long-lasting enhancement of NMDA receptor-mediated synaptic transmission by metabotropic glutamate receptor activation

John J. O'Connor<sup>\*§</sup>, Michael J. Rowan<sup>\*†</sup>  
& Roger Anwyl<sup>†</sup>

<sup>\*</sup> Department of Pharmacology and Therapeutics, and  
<sup>†</sup> Department of Physiology, Trinity College, Dublin 2, Ireland

SYNAPTIC transmission mediated by the *N*-methyl-D-aspartate (NMDA) glutamate receptor plays a key role in a range of plastic processes in the nervous system. These include long-term potentiation of synaptic transmission mediated by the  $\alpha$ -amino-3-hydroxy-5-methyl-4-isoxazolepropionate (AMPA) receptor, neuronal development, excitotoxicity and certain learning tasks<sup>1,2</sup>. Recently, long-term potentiation of NMDA receptor-mediated synaptic transmission was found to occur following high-frequency (tetanic) stimulation via an unknown mechanism<sup>3–7</sup>. We show here that activation of metabotropic glutamate (mGlu) receptors by neurally released transmitter underlies this type of long-term potentiation. The whole-cell patch-clamp technique in the 'thick' slice of the rat dentate gyrus was used to measure NMDA receptor-mediated excitatory postsynaptic currents. We have found that mGlu receptor activation by a selective agonist produced a long-lasting enhancement which was mutually exclusive with long-term potentiation of these NMDA currents. Moreover, both forms of potentiation were greatly reduced by the mGlu receptor antagonists L-2-amino-3-phosphonopropionate and (*R,S*)- $\alpha$ -methyl-4-carboxyphenylglycine.

Long-term potentiation (LTP) of the NMDA receptor-mediated excitatory postsynaptic currents (NMDA e.p.s.c.s) in response to tetanic stimulation measured  $181 \pm 16\%$  of control (mean  $\pm$  s.e.m.,  $n = 5$ ;  $P < 0.01$ ) at 20 min post-tetanus at a holding potential of  $-30$  mV, and lasted for more than 30 min without decreasing (Figs 1a and 3d). The current-voltage (*I-V*) curves (Fig. 1b) and normalized voltage-conductance (*V-g*) curves (Fig. 1c) of the NMDA e.p.s.c. before and after LTP induction demonstrated that the e.p.s.c. was enhanced at all potentials from  $-90$  to  $+10$  mV with no observable voltage-dependency. No change in the rise and decay times or the reversal potential of the NMDA e.p.s.c. occurred following tetanically induced LTP.

Activation of mGlu receptors with the selective mGlu receptor agonist aminocyclopentane-1*S*,3*R*-dicarboxylic acid (ACPD)<sup>8</sup>

evoked a long-lasting enhancement of the NMDA e.p.s.c. which was very similar to that following LTP. Perfusion of  $10 \mu\text{M}$  ACPD for 2 min caused an initial depression of the NMDA e.p.s.c. lasting 2–4 min (probably a presynaptic inhibition<sup>9,10</sup>), followed by a long-lasting enhancement measuring  $169 \pm 13\%$  at peak and  $160 \pm 13\%$  of control at 20 min after ACPD application ( $n = 8$ ;  $P < 0.01$ ; Figs 1d and 3a). *I-V* (Fig. 1e) and normalized *V-g* curves (Fig. 1f) of the NMDA e.p.s.c. before and during the enhancement by ACPD showed that its effects occurred at all potentials from  $-90$  to  $+20$  mV with no observable voltage-dependency. No change in the rise and decay time or the reversal potential of the NMDA e.p.s.c. occurred after ACPD application. Previous studies have shown that activation of the mGlu receptor induces a long-lasting increase in NMDA receptor-mediated field excitatory postsynaptic potentials (e.p.s.ps)<sup>11</sup> and reversible enhancement of responses evoked by exogenous application of NMDA<sup>12–14</sup>.

Occlusion experiments were carried out to demonstrate that the mGlu receptor-mediated enhancement shared common mechanisms of expression with LTP of the NMDA e.p.s.c. In the first set of experiments, tetanically induced LTP (with identical parameters to Fig. 1) of the NMDA e.p.s.c. was induced 5 min after a cell seal had been obtained. After a further 8 min,  $10 \mu\text{M}$  ACPD was applied. Although the initial depression of the NMDA e.p.s.c. was still observed, no further enhancement occurred (Fig. 2a); thus, the LTP of the NMDA e.p.s.c. blocked the mGlu receptor-mediated enhancement of the e.p.s.c. In the second set of experiments, ACPD ( $10 \mu\text{M}$ ) was applied to the slice 5 min after a cell seal had been obtained, resulting in the initial depression followed by an enhancement of the NMDA e.p.s.c. After a further 8 min, tetanic stimulation was applied. No LTP was induced (Fig. 2b); that is, the mGlu receptor-activated enhancement of the NMDA e.p.s.c. blocked the tetanically induced enhancement of the e.p.s.c. The potentiated synaptic currents in both occlusion experiments were completely antagonized by  $50 \mu\text{M}$  D-2-amino-5-phosphonopentanoate (AP5). The occlusion of the LTP or the ACPD-induced enhancement of the NMDA e.p.s.c. was not due to the later time of application of the tetanus of ACPD because in a separate set of control experiments both tetanically induced LTP and ACPD-induced enhancements were observed without any reduction in amplitude. These control experiments were carried out on different slices at the same time after whole-cell clamp formation (19 min) as in the occlusion experiments (see Fig. 2).

Proof of the role of the mGlu receptor in LTP of the NMDA e.p.s.c. requires the use of specific antagonists of the mGlu receptor. In the final set of experiments, we investigated the action of the mGlu receptor antagonists L-2-amino-3-phosphonopropionate (AP3)<sup>15</sup> and (*R,S*)- $\alpha$ -methyl-4-carboxyphenylglycine (MCPG)<sup>16</sup> on the ACPD enhancement and the tetanically induced LTP of the NMDA e.p.s.c. AP3 ( $200 \mu\text{M}$ ) and MCPG ( $400 \mu\text{M}$ ) both greatly reduced the amplitude and duration of the ACPD-induced enhancement of the NMDA e.p.s.c. Thus, in the presence of AP3, no ACPD-induced long-lasting enhancement of the NMDA e.p.s.c. occurred. Only a small-amplitude short-term potentiation (peak of  $128 \pm 11\%$  compared with  $169 \pm 13\%$  in control), which returned to baseline by 8 min, was present (Fig. 3b). Moreover, in MCPG, only a small-amplitude (peak  $124 \pm 11\%$ ) short-term potentiation of the NMDA e.p.s.c. occurred which returned to baseline by 15 min (Fig. 3c). Both AP3 ( $200 \mu\text{M}$ ) and MCPG ( $400 \mu\text{M}$ ) also greatly reduced the amplitude and time course of the tetanically induced potentiation, with only a short-term potentiation remaining in the presence of either of these antagonists. The peak amplitude of the potentiation following tetanic stimulation in AP3 was  $115 \pm 9\%$  of control ( $n = 5$ ; Fig. 3d, e), the potentiation returning to baseline by 7 min (Fig. 3e). The peak amplitude of the potentiation following tetanic stimulation in the presence of MCPG was  $124 \pm 11\%$ , with the potentiation lasting only 2 min (Fig. 3f).

The effects of mGlu receptor stimulation on NMDA receptor-

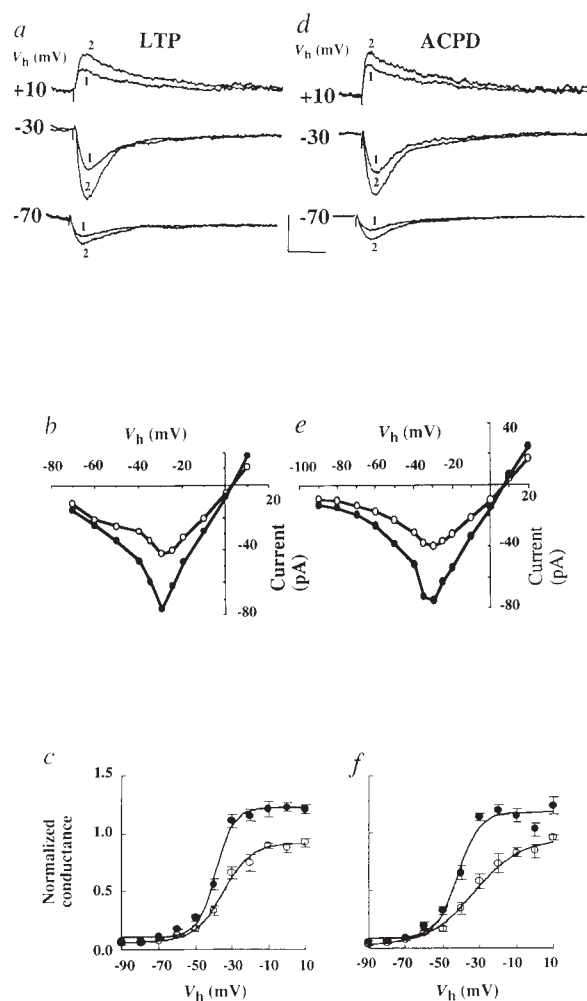
<sup>‡</sup> To whom correspondence should be addressed.

<sup>§</sup> Present address: Department of Human Anatomy and Physiology, University College Dublin, Earlsfort Terrace, Dublin 2, Ireland.



FIG. 1 Effect of tetanic stimulation and ACPD on NMDA receptor-mediated synaptic transmission. **a**, Original traces of NMDA e.p.s.cs at three different holding potentials before (1) and 15–20 min after (2) LTP induction. Each trace is the average of three individual sweeps. LTP was induced by tetanic stimulation consisting of eight trains each of eight pulses at 200 Hz, inter-train interval 2 s, under current-clamp conditions at resting potential for the duration of the tetanus. Horizontal bar, 50 ms; vertical bar, 30 pA. **b**,  $I$ - $V$  relationship of the NMDA e.p.s.c. before ( $\circ$ ) and after ( $\bullet$ ) tetanic stimulation. **c**, Normalized  $V$ - $g$  relationship of the NMDA e.p.s.cs before ( $\circ$ ) and after ( $\bullet$ ) tetanic stimulation. The peak NMDA conductance measured in the presence of 6-cyano-7-nitroquinoxaline-2,3-dione (CNQX) was calculated by the equation  $g = I/(V_h - E_0)$  where  $V_h$  is the holding potential and  $E_0$  the reversal potential. Data were fitted by nonlinear least-squares to the Boltzmann type function  $g = (g_{\max} - g_{\min}) / (1 + \exp((V' - V_h)/k)) + g_{\min}$ , where  $g_{\max}$  and  $g_{\min}$  are the maximum and minimum conductances,  $V'$  is the holding potential at which the conductance is half-maximal and  $k$  is a constant. There was a significant increase in  $g_{\max}$  and  $g_{\min}$  ( $P < 0.05$ ;  $n = 6$ ; paired  $t$ -tests). **d**, Enhancement of the NMDA receptor-mediated e.p.s.c. by 10  $\mu$ M 1S, 3R-ACPD (ACPD). Original traces of the NMDA e.p.s.cs before (1) and 10–15 min after ACPD application (2) at three different holding potentials. **e**,  $I$ - $V$  curve of the NMDA e.p.s.c. before ( $\circ$ ) and after ( $\bullet$ ) application of ACPD. **f**, Normalized  $V$ - $g$  relationship of the NMDA e.p.s.c. before ( $\circ$ ) and after ( $\bullet$ ) application of ACPD. Data were fitted as described above (see **c**). There was a significant increase in  $g_{\max}$  ( $P < 0.05$ ;  $n = 5$ ; paired  $t$ -test).

**METHODS.** Rat hippocampal slices (350  $\mu$ m thick) were prepared using standard techniques. They were submerged in a solution containing: (mM) NaCl, 120; KCl, 2.5;  $\text{NaH}_2\text{PO}_4$ , 1.25;  $\text{NaHCO}_3$ , 26;  $\text{MgSO}_4$ , 2.0;  $\text{CaCl}_2$ , 1.5; D-glucose, 10, and bubbled with 95%  $\text{O}_2$ /5%  $\text{CO}_2$  at a temperature of 32  $^\circ\text{C}$ . The patch-clamp electrode, resistance 5–8 M $\Omega$ , contained (mM): potassium gluconate, 130; KCl, 10; EGTA, 10;  $\text{CaCl}_2$ , 1;  $\text{MgCl}_2$ , 3; HEPES, 20;  $\text{Na}_2\text{ATP}$ , 3;  $\text{NaGTP}$ , 0.5; QX 314, 5; pH to 7.2 using KOH. Some experiments were carried out using caesium methanesulphonate (130 mM) in place of potassium gluconate with no resulting difference. Pharmacologically isolated (CNQX, 4–10  $\mu$ M; picrotoxin, 100  $\mu$ M) NMDA e.p.s.cs were recorded using a List EPC-7 amplifier (3 kHz low-pass Bessel filter) from rat dentate granule cells in response to stimulation of the proximal (50  $\mu$ m from the cell bodies) afferent pathway. These synaptic currents were shown to be NMDA receptor-mediated as they were completely blocked by D-2-amino-5-phosphonopentanoate. Stimulation (0.05 Hz) intensity was adjusted to evoke an NMDA e.p.s.c. which was 20–30% of the maximum amplitude (30–60 pA at a  $V_h$  of  $-30$  mV). The NMDA e.p.s.cs had very similar parameters to those found previously in hippocampal cells<sup>22,23</sup>, with a 10–90% rise time of  $7.9 \pm 2.2$  ms (mean  $\pm$  s.e.m.,  $n = 10$ ) at a holding potential of  $-30$  mV, and a biexponential decay time course with a fast and a slow decay time constant measuring  $37.5 \pm 3.5$  ms and  $343 \pm 65$  ms, respectively. The current-voltage curve of the NMDA e.p.s.c. peaked at  $-30$  mV and had a reversal potential of  $+5 \pm 3$  mV ( $n = 25$ ). Previously described procedures<sup>6,7,22–24</sup> were used to verify the adequacy of the clamping, and experimental data were only collected from cells showing



high-quality space clamping of the activated synapses. The capacitive current was always electronically cancelled and the series resistance (range 7–20 M $\Omega$ , as measured directly from the amplifier) compensated by 60–70% during measurements of synaptic currents. Mean input resistance was  $239 \pm 24$  M $\Omega$  and resting membrane potential was  $-68 \pm 4$  mV. The input resistance was monitored continuously throughout the experiments, and the recording terminated if it varied by more than 10%.

mediated excitatory synaptic transmission described here differs greatly from its effects on AMPA receptor-mediated excitatory synaptic transmission described previously<sup>16,17</sup>. Thus, ACPD generated only a slow-onset potentiation of the AMPA receptor-mediated field e.p.s.p. which reached a peak in about 90 min in the CA1 region of the hippocampus<sup>16,17</sup>. However, as shown in the present study, ACPD induced a rapid potentiation of the NMDA e.p.s.c. which peaked within about 5 min. Moreover, antagonists of the mGlu receptor were previously shown only to reduce the time course (to 30–90 min duration), and not the initial amplitude, of LTP of AMPA receptor-mediated synaptic transmission in CA1<sup>15,16</sup>. In contrast, the present study shows that mGlu receptor antagonists greatly reduce the amplitude of LTP of the NMDA e.p.s.c., including the initial potentiation shortly after tetanic stimulation. It is possible that the facilitatory effects of mGlu receptor stimulation on NMDA and AMPA receptor-mediated synaptic transmission are generated by different subclasses of mGlu receptor. Thus, the AP3-sensitive mGlu receptor responsible for the LTP of the NMDA e.p.s.c.

may be linked to the phosphoinositide system, as AP3 is known to be an antagonist at phosphoinositide-linked mGlu receptors in the hippocampus<sup>18</sup>. In contrast, the mGlu receptor linked to an increase in cyclic AMP may contribute to LTP of AMPA receptor-mediated transmission<sup>19</sup>.

An mGlu receptor-induced increase of synaptic NMDA currents is likely to be of importance in modulating the amplitude of NMDA receptor-dependent LTP of AMPA receptor-mediated synaptic transmission. It has been shown previously that the amplitude of tetanically induced LTP of AMPA receptor-mediated field e.p.s.ps was enhanced by mGlu receptor activation<sup>20,21</sup>, an enhancement produced in the presence of a small ACPD-induced depression of synaptic transmission<sup>20,21</sup>. Such enhancement is likely to have been brought about through the mGlu receptor-mediated enhancement of NMDA currents, which we have now shown is rapidly induced and overrides any transient reduction of transmission. Because of the long duration of this enhancement, it will be especially effective in modulating subsequent episodes of LTP. □

FIG. 2 Occlusion experiments. *a*, Initial tetanic stimulation (closed arrow) induced LTP of the NMDA e.p.s.c. (closed symbols). Subsequent application of  $10\ \mu\text{M}$  ACPD for 2 min (closed bar) produced a short-lasting depression, with no enhancement of the e.p.s.c. The graph shows the mean  $\pm$  s.e.m. of five experiments. Tetanic stimulation (open arrow) 19 min after obtaining whole-cell configuration (open symbols) gave rise to robust LTP. The mean of two sets of separate experiments carried out on separate slices is shown. *b*, Initial application of  $10\ \mu\text{M}$  ACPD for 2 min produced a short-lasting depression followed by an enhancement of the NMDA e.p.s.c. (closed symbols). Subsequent tetanic stimulation (closed arrow) with identical parameters to those described above failed to induce any further enhancement of the e.p.s.c. The graph shows the mean of five experiments. ACPD application (open bar) 19 min after obtaining whole-cell configuration (open symbols) produced a long-lasting enhancement. The mean of two sets of separate experiments carried out on separate slices is shown. As the potentiated responses were at most 50% of the maximum obtainable by increasing the stimulus intensity, the occlusion cannot be due to a saturation of the input-output curve.

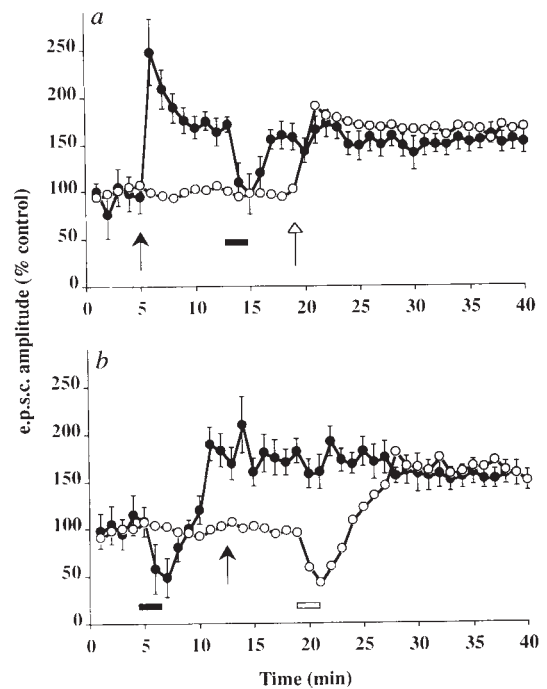
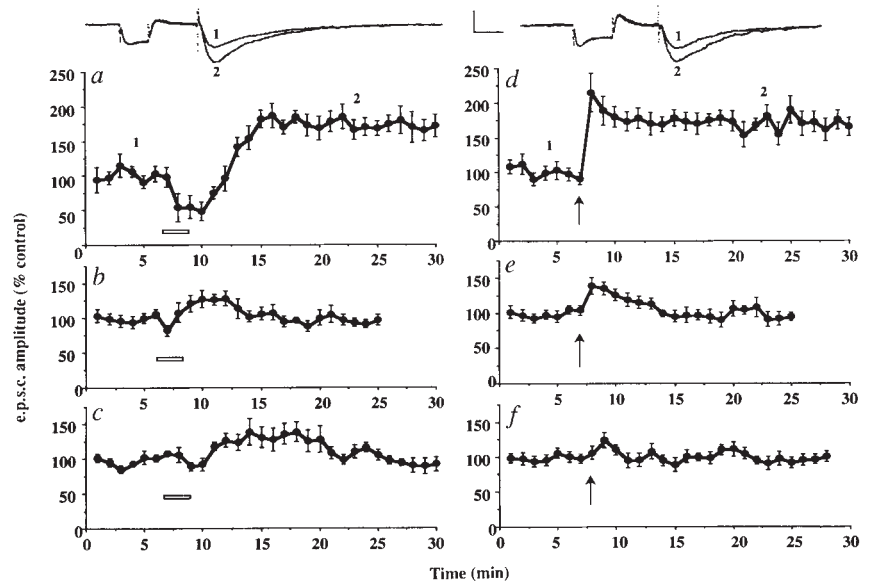


FIG. 3 mGlu receptor antagonists block the induction of LTP and ACPD-induced enhancement of the NMDA e.p.s.c. *a*, Bath perfusion of  $10\ \mu\text{M}$  ACPD for 2 min (open bar) generated an initial reduction followed by a long-lasting enhancement of the NMDA e.p.s.c. evoked at  $-30\ \text{mV}$  by step depolarization from a holding potential of  $-70\ \text{mV}$ . The graph shows the mean  $\pm$  s.e.m. of five experiments. Traces at the top show typical e.p.s.c.s 2 min before (1) and 15 min after (2) ACPD application. Each stimulus was preceded by a  $20\ \text{mV}$  hyperpolarizing pulse. There was no change in the input conductance. Calibration: vertical bar,  $30\ \text{pA}$ ; horizontal bar,  $50\ \text{ms}$ . *b*, Bath perfusion of  $200\ \mu\text{M}$  AP3 for at least 1 h before ACPD application. The ACPD-induced long-lasting enhancement of the NMDA e.p.s.c. was significantly reduced in amplitude and time course ( $P < 0.01$ ; Mann-Whitney  $U$ ). The graph shows the mean of five experiments. *c*, Bath perfusion of  $400\ \mu\text{M}$  MCPG for at least 1 h before ACPD application. The ACPD-induced long-lasting enhancement of the NMDA e.p.s.c. was significantly reduced in amplitude and duration ( $P < 0.01$ ; Mann-Whitney  $U$ ). The graph shows the mean of four experiments. *d*, Tetanic stimulation with the same parameters as in Fig. 1 (arrow) resulted in a rapid and long-lasting increase in the peak amplitude of the e.p.s.c. evoked at  $-30\ \text{mV}$  (applied by voltage steps from a holding potential of  $-70\ \text{mV}$ ). The graph shows the mean of five experiments. Traces show typical e.p.s.c.s 2 min before (1) and 15 min after (2) tetanic stimulation. *e*, Bath perfusion of  $200\ \mu\text{M}$  AP3 for at least 1 h before tetanic



stimulation. The LTP of the NMDA e.p.s.c. was significantly reduced in amplitude and duration ( $P < 0.01$ ; Mann-Whitney  $U$ ). *f*, Bath perfusion of  $400\ \mu\text{M}$  MCPG for at least 1 h before tetanic stimulation. The LTP was completely blocked ( $P < 0.01$ ; Mann-Whitney  $U$ ), leaving only a very small amplitude and brief-duration short-term potentiation. The graph shows the mean of four experiments.

Received 17 May; accepted 3 December 1993.

- Collingridge, G. L. & Lester, R. A. *Pharmac. Rev.* **40**, 143–210 (1989).
- Bliss, T. V. L. & Collingridge, G. L. *Nature* **361**, 31–39 (1993).
- Asztely, F., Wigstrom, H. & Gustafsson, B. *Eur. J. Neurosci.* **4**, 681–690 (1992).
- Bashir, Z., Alford, S., Davies, S. N., Randall, A. D. & Collingridge, G. L. *Nature* **349**, 156–158 (1991).
- Beretta, N. et al. *Eur. J. Neurosci.* **3**, 850–854 (1991).
- Xie, X., Berger, T. W. & Barrionuevo, G. *J. Neurophysiol.* **67**, 1009–1013 (1992).
- Muller, D., Arai, A. & Lynch, G. *Hippocampus* **2**, 29–38 (1992).
- Palmer, E., Monaghan, D. T. & Cotman, C. W. *Eur. J. Pharmac.* **166**, 585–587 (1989).
- Baskys, A. & Malenka, R. C. *J. Physiol., Lond.* **444**, 687–701 (1991).
- Lovinger, D. M. *Neurosci. Lett.* **129**, 17–21 (1991).
- Collins, C. G. S. *Br. J. Pharmac.* **108**, 422–430 (1993).
- Aniksztejn, L., Otani, S. & Ben-Ari, Y. *Eur. J. Neurosci.* **4**, 500–505 (1992).
- Cerne, R. & Randic, M. *Neurosci. Lett.* **144**, 180–184 (1992).

- Collingridge, G. L. et al. *Int. Acad. Biomed. Drug Res.* **2**, 41–49 (1992).
- Behnisch, T., Fjodorov, K. & Reymann, K. G. *NeuroReport* **2**, 386–388 (1991).
- Bashir, Z. I. et al. *Nature* **363**, 347–350 (1993).
- Bortolotto, S. A. & Collingridge, G. L. *Neuropharmacology* **32**, 1–9 (1993).
- Nicoletti, F. et al. *J. Neurochem.* **46**, 40–46 (1986).
- Musgrave, M. A., Ballyk, B. A. & Goh, J. W. *NeuroReport* **4**, 171–174 (1993).
- McGuinness, N., Anwyl, R. & Rowan, M. J. *Eur. J. Pharmac.* **197**, 231–232 (1991).
- McGuinness, N., Anwyl, R. & Rowan, M. J. *NeuroReport* **2**, 688–690 (1991).
- Hestrin, S., Nicoll, R. A., Perkel, D. J. & Sah, P. J. *Physiol., Lond.* **422**, 203–225 (1990).
- Keller, B. U., Konnerth, A. & Yaari, Y. *J. Physiol., Lond.* **435**, 275–293 (1991).
- Llano, I., Marty, A., Armstrong, C. M. & Konnerth, A. *J. Physiol., Lond.* **434**, 183–213 (1991).

ACKNOWLEDGEMENTS. The software for the analysis of synaptic currents was provided by J. Dempster. We thank S. Logan, H. Wheal and G. L. Collingridge for assistance with the patch-clamp technique. We thank the Health Research Board, Ireland, and the Wellcome Trust for financial support.



# S proteins control rejection of incompatible pollen in *Petunia inflata*

Hyun-Sook Lee, Shihshieh Huang & Teh-hui Kao\*

Department of Biochemistry and Molecular Biology, and Center for Gene Regulation, Pennsylvania State University, University Park, Pennsylvania 16802, USA

FLOWERING plants have evolved various stratagems to prevent inbreeding and promote outcrosses<sup>1</sup>. One such mechanism, gametophytic self-incompatibility, provides a genetic barrier to self-fertilization, and in the simplest cases is controlled by the highly polymorphic *S* locus<sup>2</sup>. Growth of a pollen tube in the style is arrested when the *S* allele carried by the pollen matches one of the two *S* alleles carried by the pistil. Putative *S* allele proteins of the pistil have been identified in several solanaceous species based on their co-segregation with *S* alleles<sup>3–12</sup>, and they have been shown to be ribonucleases<sup>13–15</sup>. So far, there has been only correlative or indirect evidence for the claim that these *S* allele-associated proteins (*S* proteins) are involved in recognition and rejection of self pollen<sup>16,17</sup>. Here we show that inhibition of synthesis of *S*3 and *S*2

proteins in *Petunia inflata* plants of *S*2*S*3 genotype by the antisense *S*3 gene resulted in failure of the transgenic plants to reject *S*3 and *S*2 pollen. We further show that expression of the transgene encoding *S*3 protein in *P. inflata* plants of *S*1*S*2 genotype confers on the transgenic plants the ability to reject *S*3 pollen. The self-incompatibility behaviour of the pollen was not affected by the transgene in either set of experiments. Taken together, these findings provide direct *in vivo* evidence that *S* proteins control the self-incompatibility behaviour of the pistil.

We used both loss-of-function and gain-of-function approaches to ascertain whether the *P. inflata* *S* proteins we previously identified<sup>9</sup> do indeed control self-incompatibility interactions between pollen and pistil. In the former approach, we introduced antisense *S*3 complementary DNA under the control of the *S*3 gene promoter into *P. inflata* plants of *S*2*S*3 genotype by *Agrobacterium*-mediated transformation (Fig. 1a). This allowed us to determine whether inhibition of *S*3 protein synthesis in the pistil abolished the ability of the transgenic plants to reject *S*3 pollen.

Forty-seven independent transgenic plants were self-pollinated to determine whether their self-incompatibility behaviour had been affected. Of 30 plants found to set variable numbers of seeds, six (AS-4, AS-14, AS-23, AS-27, AS-35 and AS-37) consistently produced a number of seeds comparable to that obtained from compatible pollination of *P. inflata* plants (~200 per fruit), and these were chosen for further analysis. Genomic

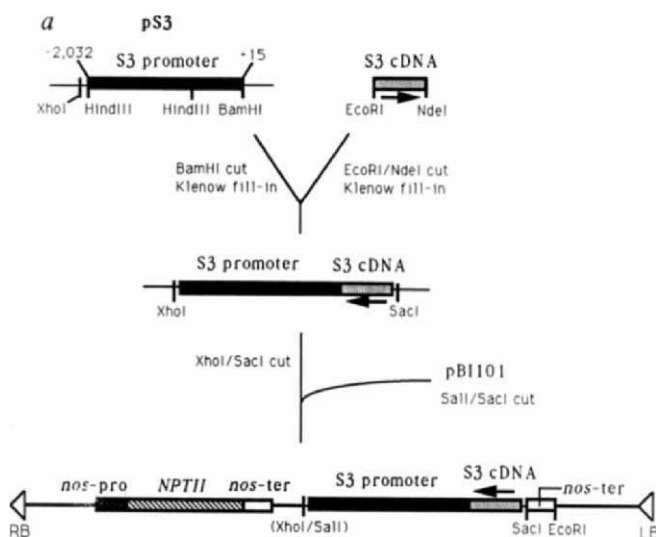


FIG. 1 Transformation of *P. inflata* plants with antisense *S*3 gene and analysis of transgenic plants for presence of the transgene. **a**, Antisense construct. The promoter used to express the antisense *S*3 cDNA was contained in a DNA fragment spanning  $-2,032$  to  $+15$  base pairs (bp) of the *S*3 gene<sup>18</sup>. This DNA fragment confers pistil expression of the gene encoding  $\beta$ -glucuronidase (GUS) in transgenic *P. inflata* plants (our unpublished results). The *EcoRI*–*NdeI* fragment of the *S*3 cDNA used in the construct contained  $\sim 70\%$  of the full-length *S*3 cDNA previously reported<sup>9</sup>. Sense *S*3 cDNA is indicated by an arrow pointing right; antisense *S*3 cDNA is indicated by an arrow pointing left. The transcriptional termination signal was provided by the nopaline synthase terminator (*nos-ter*) present on a binary Ti plasmid pBI101 (Clontech). The *NPTII* gene, which encodes neomycin phosphotransferase and confers kanamycin resistance in transgenic plants, is expressed by the *nos* gene promoter (*nos-pro*). RB and LB denote right and left borders of T-DNA. **b**, Genomic Southern blot. The filter containing *EcoRI* digests of genomic DNA was hybridized to a probe containing full-length *S*3 cDNA<sup>9</sup>. Lane 1, parental *S*2*S*3 plant; lane 2, AS-37; lane 3, AS-4; lane 4, AS-14; lane 5, AS-27. The 11.5-kb fragment present in both the parental *S*2*S*3 plant and the four transgenic plants corresponds to the endogenous *S*3 gene. A 2.3-kb fragment corresponding to the *S*2 gene is not shown. The 21-

and 5.1-kb DNA size markers are indicated. Additional fragments seen in the transgenic plants, but not in the parental *S*2*S*3 plant, each resulted from one cut by *EcoRI* in the integrated antisense *S*3 gene and one cut in the genome. AS-27 and AS-37 contained one insertion, AS-4 two insertions and AS-14 three insertions. Multiple insertions of the transgene in AS-4 and AS-14 were also confirmed by *HindIII* digestions (results not shown).

**METHODS.** **a**, The 2,047-bp *S*3-promoter fragment was cloned into the *HindIII* and *SmaI* sites of pBluescript KS+ (Stratagene) to yield pS3. pS3 was digested with *BamHI* and Klenow enzyme was used to create blunt ends. Similarly, blunt ends were created on an *EcoRI*–*NdeI* fragment of *S*3 cDNA, and this fragment was ligated in antisense orientation to the *S*3 promoter in pS3. The *S*3-promoter–*S*3-cDNA fusion product was released by double digestion with *XhoI* and *SacI*, and the fragment was ligated into the *SacI* and *SacI* sites of pBI101. (The GUS gene present on pBI101 was removed during this step.) The recombinant Ti plasmid, designated pAS3, was electroporated into *Agrobacterium tumefaciens* strain LBA4404. Leaf disks of *P. inflata* with *S*2*S*3 genotype were infected with the *Agrobacterium* by the co-cultivation method<sup>21</sup> on MS medium supplemented with benzylaminopurine ( $1.0 \text{ mg l}^{-1}$ ) and naphthalene acetic acid ( $75 \text{ } \mu\text{g l}^{-1}$ ). Shoots were regenerated on fresh MS medium<sup>22</sup> supplemented with kanamycin ( $100 \text{ } \mu\text{g ml}^{-1}$ ) and carbenicillin ( $500 \text{ } \mu\text{g ml}^{-1}$ ). Regenerated shoots were transferred to hormone-free MS medium containing the same concentrations of antibiotics to induce root formation. **b**, Genomic DNA was isolated from 2 g of frozen leaves with an Elu-Quick kit (Schleicher & Schuell) following the manufacturer's protocol. Genomic DNA ( $5 \text{ } \mu\text{g}$ ) was digested with *EcoRI*, separated on a 0.8% agarose gel and transferred to a Biotrans (+) nylon membrane (ICN). The membrane was prehybridized in a solution containing  $5 \times \text{SSC}$ ,  $5 \times \text{Denhardt's}$ ,  $0.1\%$  SDS and  $100 \text{ } \mu\text{g ml}^{-1}$  of sheared and denatured fish sperm DNA for 2 h at  $65^\circ\text{C}$ . Hybridization was carried out in the same solution with the addition of  $^{32}\text{P}$ -labelled *S*3 cDNA for 16 h at  $65^\circ\text{C}$ . The membrane was washed in  $0.1 \times \text{SSC}$  and  $0.1\%$  SDS at  $65^\circ\text{C}$  for 1 h and exposed on X-ray film with an intensifying screen for 2 days at  $-70^\circ\text{C}$ .

\* To whom correspondence should be addressed

blot analysis of four of the plants revealed that, in addition to the 11.5-kilobase (kb) DNA fragment corresponding to the endogenous *S3* gene, they contained one to three insertions of the transgene (Fig. 1b).

To determine whether the breakdown of self-incompatibility in the six self-compatible transgenic plants resulted from loss of their ability to reject *S3* pollen, these plants were pollinated with pollen from *S2S3* and *S2S2* tester plants. AS-14, AS-23 and AS-37 rejected pollen from *S2S2* plants, but accepted pollen from *S2S3* plants, indicating that the *S3* allele had been rendered non-functional, but that the *S2* allele was not affected. AS-4, AS-27 and AS-35 accepted pollen from both *S2S2* and *S2S3* plants, indicating that, in addition to the *S3* allele, the *S2* allele had also been rendered non-functional. AS-39, a self-incompatible

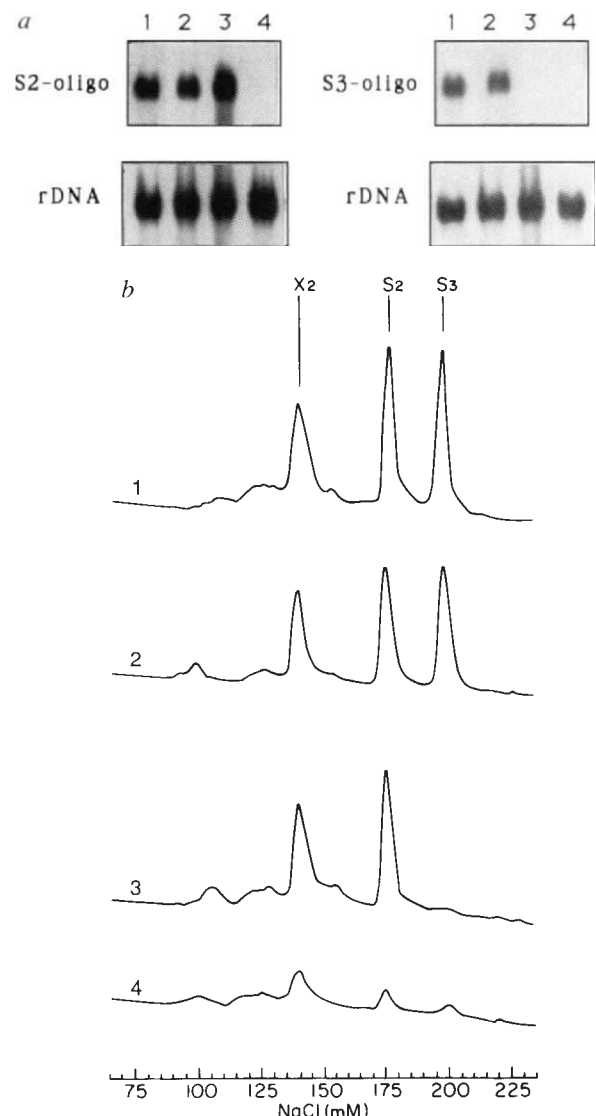
transgenic plant, rejected pollen from both *S2S2* and *S2S3* plants.

We next investigated whether the inability of these six plants to reject *S2* or *S3* pollen was caused by inhibition of *S2*- or *S3*-protein synthesis. As shown in Fig. 2a, AS-14, which rejected the *S2* but not the *S3* allele, contained normal levels of *S2* RNA, but undetectable levels of *S3* RNA; AS-27, which rejected neither the *S2* nor *S3* allele, contained undetectable levels of *S2* RNA and *S3* RNA; AS-39, which rejected both *S2* and *S3* alleles, contained normal levels of *S2* and *S3* RNA.

We then determined the amounts of *S2* and *S3* proteins in the pistils of the six self-compatible and seven self-incompatible transgenic plants, and in seven parental *S2S3* plants. Profile 1 (Fig. 2b) is representative of the parental *S2S3* plants. AS-14,

FIG. 2 Analysis of the amounts of *S2* and *S3* RNA, and *S2* and *S3* proteins in a parental *S2S3* plant and transgenic plants. a, Northern blots. Two identical RNA blots, each containing total pistil RNA (10 µg per lane) of a parental *S2S3* plant (lane 1), AS-39 (lane 2), AS-14 (lane 3) and AS-27 (lane 4), were hybridized with two radiolabelled probes separately: *S2*-oligo, an oligonucleotide specific for sense *S2* RNA; *S3*-oligo, an oligonucleotide specific for sense *S3* RNA. After autoradiography, the bound radiolabelled probes were removed from the blots, which were hybridized with the third probe, ribosomal DNA, encoding 25S rRNA of *P. inflata* (J. Mu and T.-h. K., unpublished results). b, Fast protein liquid chromatography (FPLC) profiles. *S2* and *S3* proteins cannot be separated by SDS-PAGE because of their similar molecular weights<sup>9</sup>, but can be separated from each other and from other pistil proteins by cation-exchange chromatography on a Mono-S column using the FPLC system (Pharmacia)<sup>14</sup>. Most pistil proteins flowed through during loading and washing of the column with 50 mM sodium phosphate (pH 6.0). The bound proteins that were eluted with the salt gradient consisted mainly of S proteins and a pistil-specific ribonuclease, X2 (refs 14, 23). Profile 1, parental *S2S3* plant; 2, AS-39; 3, AS-14; 4, AS-27. The observed concomitant reduction of the level of *S2* protein in AS-27 (as well as in AS-4 and AS-35, not shown) and of the level of X2 in AS-27 by the antisense *S3* gene may be due to the sequence similarity between the 585-bp fragment of *S3* cDNA used in the antisense construct and the corresponding regions in the genes for *S2* protein<sup>9</sup> and RNase X2<sup>23</sup>. It is known that antisense RNA can inhibit the messenger RNA production from a target gene, as well as from genes with sequence similarity to the target gene<sup>24</sup>.

**METHODS.** a, Total RNA was isolated from pistils as described previously<sup>10</sup>. RNA samples were electrophoresed on 1.2% agarose/formaldehyde gels and transferred to Biotrans (+) membranes. The probe specific to sense *S2* RNA was a 51-mer with sequences: 5'-CAGAAC-ATTGATTATATTATCTTCTTTAAACGCGAATACTTGTCCAGT-3'. The sequence is complementary to a segment of *S2* cDNA encoding amino acids 48–64 of *S2* protein<sup>9</sup> located in the hypervariable region HVa of S proteins<sup>25</sup>. The probe specific to sense *S3* RNA was a 54-mer with sequence: 5'-CAAATCATTGACAAITCTATCTTTAAGCTGAACGACAAAC-TTATCTCCATC-3'. This sequence is complementary to a segment of *S3* cDNA encoding amino acids 48–65 of *S3* protein<sup>9</sup>, also located in the HVa region. The two oligonucleotides were <sup>32</sup>P-labelled at their 5' ends with T4 polynucleotide kinase. For hybridization with *S2*- or *S3*-oligo, the membranes were prehybridized as described for Fig. 1b, except at 45 °C for 2 h, and hybridized in the prehybridization solution containing *S2*- or *S3*-oligo probe at 45 °C overnight. The membranes were washed twice in 0.1 × SSC, 0.1% SDS at room temperature for 10 min each, and then washed with the same solution at 40 °C for 5 min. Autoradiography was carried out at –70 °C for 16 h with an intensifying screen. The bound radiolabelled probes were removed from the membranes by boiling in 0.1 × SSC and 0.1% SDS. For hybridization with the rDNA probe, the membranes were prehybridized as described for Fig. 1b and hybridized with the prehybridization solution containing <sup>32</sup>P-labelled rDNA probe at 65 °C overnight. They were washed in 0.1 × SSC, 0.1% SDS at 65 °C for 1 h and autoradiographed at –70 °C for 1 h with an intensifying screen. b, Pistils were collected from freshly opened flowers of each plant and stored at –70 °C until use, when 40 mg from each plant was ground to a fine powder in liquid nitrogen, with further grinding after the addition of 1 ml of the extraction buffer containing 50 mM Tris-HCl, pH 8.5, 10 mM EDTA, 1 mM phenylmethylsulphonyl fluoride, 1 mM CaCl<sub>2</sub> and 1 mM dithiothreitol. The crude extract was centrifuged



at 12,000g for 10 min, and the supernatant filtered through a 0.45-µm Millex-GV filter (Millipore) to remove unsedimented fine particles. The filtrate was applied to a Mono-S column (HR 5/5) equilibrated with 50 mM sodium phosphate (pH 6.0). The bound proteins were eluted with a linear gradient of 0–500 mM NaCl in the same buffer at a flow rate of 0.5 ml min<sup>–1</sup>. Proteins were monitored at A<sub>280</sub> with the sensitivity of the detector set to 0.1 AUFS. No discernible differences in the FPLC profiles for each plant were detected when two separated pistil extractions were used. The profile shown for each plant is one of the two runs.



AS-23 and AS-37 all contained normal amounts of S2 protein but drastically reduced amounts of S3 protein (profile 3 shows AS-14). Levels of both S2 and S3 proteins were drastically reduced in AS-4, AS-27 and AS-35 (profile 4 shows AS-27). In fact, the amounts of S3 protein in the former three plants, and the amounts of S2 and S3 in the latter three, were lower than those present in immature buds which are fully receptive to self pollen<sup>9</sup>. Amounts of S2 and S3 proteins of the seven self-incompatible transgenic plants were comparable to those of the parental S2S3 plants (Fig. 2b, profile 2, shows AS-39).

The effect of the antisense S3 gene was inheritable because

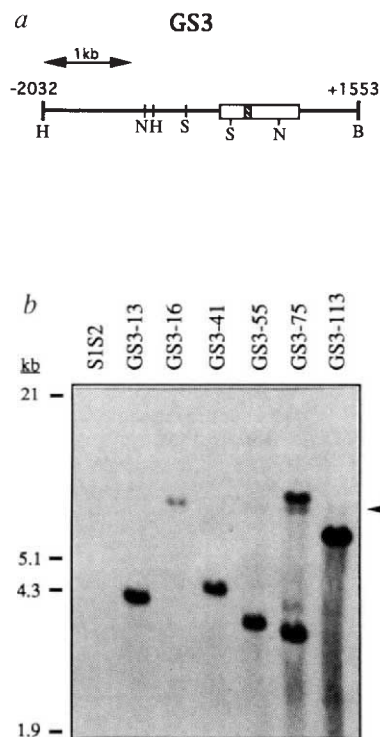


FIG. 3 Transformation of *P. inflata* plants with the S3 gene and analysis of transgenic plants for presence of the transgene. *a*, Schematic representation of the genomic DNA used in transformation experiments. The DNA fragment, GS3, spans  $-2,032$  to  $+1,553$  bp of the S3 gene<sup>18</sup>. Open boxes denote the exons; the hatched box denotes the intron. Restriction enzyme sites are indicated with one-letter abbreviations: H, *Hind*III; N, *Nde*I; S, *Stu*I; B, *Bsu*96I. *b*, Genomic Southern blot. The filter containing *Hind*III digests of genomic DNA was hybridized with a radiolabelled probe of the full-length S3 cDNA<sup>9</sup>. The genomic DNA was isolated from one non-transgenic S1S2 plant and six transgenic plants, as indicated above the blot. The arrow marks a weakly hybridizing DNA fragment of 9.9 kb which corresponds to the endogenous S2 gene. The endogenous S1 gene did not cross-hybridize with the S3 cDNA probe under the conditions used. The DNA size markers are indicated. Two fragments seen in GS3-75 each resulted from one cut by *Hind*III within the integrated S3 gene and a second cut in the genome.

METHODS. *a*, The 3.6-kb genomic DNA (GS3) was released from the cloning vector pBluescript KS+ (Stratagene) by digestion with *Xho*I and *Sac*I, and the fragment was ligated into *Sac*I and *Sac*I sites of a binary Ti-plasmid vector pBI101. The GUS gene present on pBI101 was removed during this step. The recombinant Ti plasmid was introduced into *P. inflata* plants by *Agrobacterium*-mediated transformation as described for Fig. 1a. *b*, Genomic DNA was isolated from 2 g of frozen leaves as described for Fig. 1b. *b*, Genomic DNA (5  $\mu$ g) was digested with *Hind*III, separated on a 0.8% agarose gel and transferred to a Biotrans (+) nylon membrane (ICN). Prehybridization, hybridization and washing of the filter were carried out as described for Fig. 1b. The filter was exposed on X-ray film at  $-70^{\circ}\text{C}$  for 48 h with an intensifying screen.

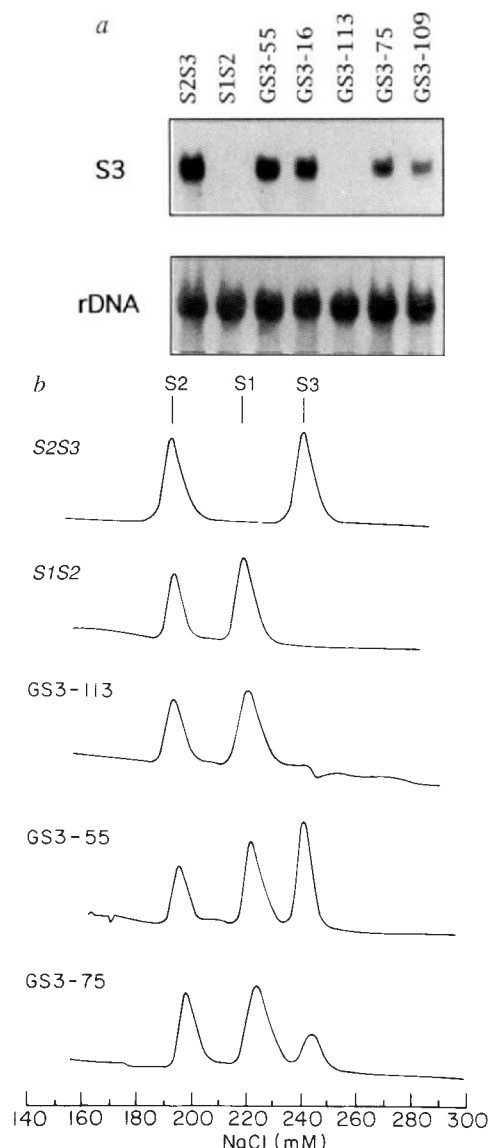


FIG. 4 Analysis of the amounts of S3 RNA and S3 protein. *a*, Northern blot. Each lane of the RNA blot contains 10  $\mu$ g of total pistil RNA. Of the seven plants analysed, S2S3 and S1S2 are non-transgenic plants and the other five are transgenic plants. The blot was first hybridized with a radiolabelled oligonucleotide probe (S3) specific to S3 RNA. After autoradiography, the bound radiolabelled probe was removed and the blot was hybridized with a second probe (rDNA) encoding 25S rRNA of *P. inflata*. The absence of a hybridizing fragment in the S1S2 sample confirms the specificity of the S3 probe used. *b*, FPLC profiles. Total pistil protein of each plant was separated on a Mono-S column. Only the S protein fractions are shown.

METHODS. *a*, The procedures for isolation of total RNA and for northern analysis were identical to those described for Fig. 2a, as were the S3-RNA-specific probe and the rDNA probe used. After washing, the amount of radioactivity associated with each hybridizing band was determined using a Betagen Betascope. The amount of S3 RNA in each transgenic plant relative to that of the S2S3 plant was calculated after correction for the differences in the total amount of rRNA. The filters were then exposed on X-ray films at  $-70^{\circ}\text{C}$  for 30 min (rDNA probe) and 5 h (S3 probe) with intensifying screens. *b*, Extractions of total pistil protein and conditions for column chromatography were as described for Fig. 2b, except that: 5 mg of the pistils from each plant were used for extractions; the gradient was twice as shallow; proteins were monitored with the sensitivity of the detector set to 0.02 AUFS; and a different Mono-S column with the same dimension was used. Some of these modifications may account for differences in the salt concentration at which the same S protein was eluted in the profiles shown here and in Fig. 2b.

self-compatible plants were found in self-fertilized progeny of AS-14 and AS-27. As in the parental transgenic plants, amounts of S2 and S3 proteins also correlated perfectly with breeding behaviour (results not shown). Thus, we conclude from the antisense experiments that S proteins are necessary for the pistil to reject self pollen.

We then used the gain-of-function approach to ascertain whether S proteins alone are sufficient for the pistil to reject self pollen. We introduced a 3.6-kb DNA fragment (Fig. 3a) containing the gene encoding S3 protein<sup>9,18</sup> into *SIS2 P. inflata* plants and determined whether expression of the S3 gene could confer on the transgenic plants a new S3 allele specificity. Eighty-one transgenic plants were pollinated with pollen from S3S3 plants, and four, GS3-13, GS3-16, GS3-41 and GS3-55, were found to reject S3 pollen completely. The rest set fruits of variable sizes with seed numbers ranging from 20 to 200, indicating that they were either partially or fully compatible with S3 pollen.

The four transgenic plants that completely rejected S3 pollen also rejected pollen from *SIS2* plants. However, they set large fruits when self-fertilized at the immature-bud stage, when self-incompatibility is not yet expressed. Thus, their failure to set fruits when pollinated with S1, S2 and S3 pollen was a true self-incompatible response, and not due to female sterility resulting from tissue culture manipulations.

Genomic blot analysis revealed that all four plants contained a single insert of the transgene (Fig. 3b). Two other transgenic plants were also included in the analysis: GS3-75, partially compatible with S3 pollen (producing ~50 seeds per fruit), and GS3-113, fully compatible with S3 pollen (producing ~200 seeds per fruit). GS3-75 contained two inserts and GS3-113 one (Fig. 3b).

To investigate whether the new S3 allele specificity acquired by the four transgenic plants resulted from expression of the S3 transgene in the pistil, we first examined the level of S3 RNA in the pistils of representative transgenic plants. GS3-16 and GS3-55, which completely rejected S3 pollen, contained normal levels of S3 RNA; GS3-75 and GS3-109, which were partially compatible with S3 pollen, contained ~30% of the normal level of S3 RNA; GS3-113, which was fully compatible with S3 pollen, contained no detectable S3 RNA (Fig. 4a).

We next examined the amounts of S3 protein in the pistils of 31 transgenic plants, including all those described above (Fig. 4b). All four plants that completely rejected S3 pollen produced normal levels of S3 protein, in addition to producing normal levels of S1 and S2 proteins (Fig. 4b; compare the profile of GS3-55 with those of *SIS2* and *S2S3* plants). All 17 plants that were partially compatible with S3 pollen consistently produced much lower than normal amounts of S3 protein (Fig. 4b; compare the profiles of GS3-75 and GS3-55) which were comparable to or less than those present in immature buds of non-transgenic plants. None of the 10 plants fully compatible with S3 pollen produced any detectable S3 protein (see the profile of GS3-113). Thus, we conclude that the production of a normal level of S3 protein in the pistil of the four transgenic plants conferred on them the ability to reject S3 pollen completely. The finding that low-level expression of the S3 transgene in the pistil of mature flowers was insufficient for rejection of S3 pollen is similar to previous findings that low-level S gene expression in immature *Petunia* buds correlated with their inability to reject self pollen<sup>9,19</sup>. Neither the S3 transgene nor the antisense S3 gene affected the self-incompatibility behaviour of the pollen of the transgenic plants (data not shown), results consistent with the model of the S locus in which separate but closely linked genes control the self-incompatibility phenotype of the pollen and pistil<sup>20</sup>.

In conclusion, our results provide direct *in vivo* evidence that the S proteins of *P. inflata* are necessary and sufficient for the pistil to reject self pollen. This study also demonstrates the feasibility of using the antisense RNA approach to break down self-incompatibility. Further, demonstration that the S phenotype of a self-incompatible plant can be altered by the introduction of

a single gene, the S gene, should allow future dissection of the functional domains of the S protein through mutagenesis studies. □

Received 12 July; accepted 10 December 1993.

- Frankel, R. & Galun, E. in *Monographs on Theoretical and Applied Genetics* (eds Frankel, R., Gali, G. A. E. & Linskens, H. F.) 5–50 (Springer, Berlin, 1977).
- de Nettancourt, D. in *Monographs on Theoretical and Applied Genetics* (eds Frankel, R., Gali, G. A. E. & Linskens, H. F.) 28–57 (Springer, Berlin, 1977).
- Bredemeijer, G. M. M. & Blass, J. *Theor. appl. Genet.* **59**, 185–190 (1981).
- Anderson, M. A. *et al. Nature* **321**, 38–44 (1986).
- Kamboj, R. K. & Jackson, J. F. *Theor. appl. Genet.* **71**, 815–819 (1986).
- Mau, S. L. *et al. Planta* **169**, 184–191 (1986).
- Broothaerts, W. J. *et al. Pl. molec. Biol.* **14**, 93–102 (1989).
- Kirch, H.-H., Uhrig, H., Lottspeich, F., Salamini, F. & Thompson, R. D. *Theor. appl. Genet.* **78**, 581–588 (1989).
- Al, Y. *et al. Sex. Pl. Reprod.* **3**, 130–138 (1990).
- Kheyr-Pour, A. *et al. Sex. Pl. Reprod.* **3**, 88–97 (1990).
- Xu, B., Grun, P., Kheyr-Pour, A. & Kao, T.-h. *Sex. Pl. Reprod.* **3**, 54–60 (1990).
- Al, Y., Kron, E. & Kao, T.-h. *Molec. gen. Genet.* **230**, 353–358 (1991).
- McClure, B. A. *et al. Nature* **342**, 955–957 (1989).
- Singh, A., Al, Y. & Kao, T.-h. *Pl. Physiol.* **96**, 61–68 (1991).
- Broothaerts, W., Vanvinckenroye, P., Decock, B., Van Damme, J. & Vendrig, J. C. *Sex. Pl. Reprod.* **4**, 258–266 (1991).
- Singh, A. & Kao, T.-h. in *Sexual Reproduction in Flowering Plants* (eds Russell, S. D. & Dumas, C.) 449–483 (Academic, San Diego, 1992).
- Sims, T. L. *Crit. Rev. Pl. Sci.* **12**, 129–167 (1993).
- Coleman, C. E. & Kao, T.-h. *Pl. molec. Biol.* **18**, 725–737 (1992).
- Clark, K. R., Okuley, J. J., Collins, P. D. & Sims, T. L. *Pl. Cell* **2**, 815–826 (1990).
- Lewis, D. *Heredity* **3**, 339–355 (1949).
- Horsch, R. B. *et al. Science* **227**, 1229–1231 (1985).
- Linsmaier, E. M. & Skoog, F. *Physiol. Pl.* **18**, 100–127 (1965).
- Lee, H.-S., Singh, A. & Kao, T.-h. *Pl. molec. Biol.* **20**, 1131–1141 (1992).
- Oeller, P. W., Min-Wong, L., Taylor, L. P., Pike, D. A. & Theologis, A. *Science* **254**, 437–439 (1991).
- Ioerger, T. R., Gohlke, J. R., Xu, B. & Kao, T.-h. *Sex. Pl. Reprod.* **4**, 81–87 (1991).

ACKNOWLEDGEMENTS. We thank C. Coleman and A. Singh for establishing the conditions for transformation and regeneration of *P. inflata*, D. Tu for use of his FPLC system, H. Flores for use of a tissue culture facility, J. Marsicano and M. Gearing for helping with greenhouse work, and A. Clark, D. Cosgrove, R. Cyr, A. Richman, S. Schaeffer and A. Singh for comments. H.-S. L. was partially supported by an NSF postdoctoral fellowship in plant biology. This work was supported by grants from the NSF and USDA to T.-h. K.

## S-RNase expressed in transgenic *Nicotiana* causes S-allele-specific pollen rejection

Jane Murfett\*, Tammy L. Atherton\*,  
Beiquan Mou\*, Charles S. Gasser†  
& Bruce A. McClure\*†

\* Department of Biochemistry, University of Missouri–Columbia,  
117 Schweitzer Hall, Columbia, Missouri 65211, USA  
† Section of Molecular and Cellular Biology,  
University of California–Davis, Davis, California 95616, USA

MANY angiosperms employ self-incompatibility systems to prevent inbreeding<sup>1,2</sup>. The simple genetics of such systems<sup>3–6</sup> have made them attractive models of plant cellular communication. Implicit in the single locus genetics is that only one or a few gene products are necessary for recognition and rejection of incompatible pollen. Results in the sporophytic system of the Brassicaceae suggest that different S-locus products are responsible for the pollen and pistil parts of the recognition and rejection response<sup>7,8</sup>. In solanaceous plants, which have a gametophytic self-incompatibility system, the S-locus product responsible for the pollen portion of the interaction has not been identified, but ribonucleases encoded by the S-locus (S-RNases) are strongly implicated in the style part of the recognition and rejection reaction<sup>9–14</sup>. In *Nicotiana glauca*, pollen recognition and rejection occur if its S-allele matches either S-allele in the style. The putative stylar S-RNase is abundant in the transmitting

† To whom correspondence should be addressed.



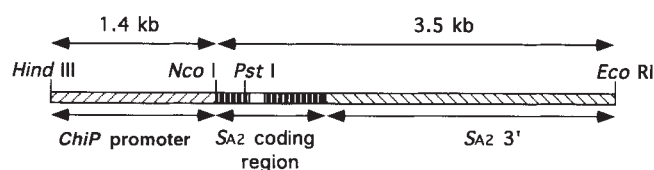


FIG. 1 Construction of a chimaeric  $S_{A2}$ -RNase gene. The  $S_{A2}$ -RNase gene sequences were carried on a 3.5-kb  $Nco$ I- $Eco$ RI fragment. The  $Nco$ I site contains the ATG translation initiation codon of the  $S_{A2}$ -RNase pre-protein. A small intron interrupts the coding region (unshaded area). The tomato  $ChiP$  promoter is carried on a 1.4-kb  $Hind$ III- $Nco$ I fragment. Fusion of the two gene fragments results in a chimaeric  $ChiP$ - $S_{A2}$ -RNase gene. The expected transcript contains 5' untranslated sequences derived from  $ChiP$ , and should produce a pre-protein identical to the precursor of authentic  $S_{A2}$ -RNase.

**METHODS.** The  $S_{A2}$ -allele was identified, characterized and cloned according to established procedures<sup>14,18-20,25,26</sup>. The  $S_{A2}$ -RNase coding region was cloned on a 6.2-kb genomic  $Eco$ RI fragment containing a single  $Pst$ I site 218 bp downstream of the ATG. A mutagenic oligonucleotide, 5'-AGCCAAGCTTACCATGGTTAATTCACAGATTACGTCAGCTG-3', and an oligonucleotide complementary to a sequence downstream of the  $Pst$ I site, 5'-CCGAATTCGAGCTCGTGCAAGTTTACATCTTTTCGG-3', were used in a PCR to amplify a region containing the N-terminal portion of the  $S_{A2}$ -RNase gene in which the translation initiation site was mutated to an  $Nco$ I site. The PCR product was inserted into the  $Pst$ I site of the genomic clone to allow isolation of the coding region, intron and downstream sequences on the 3.5-kb  $Nco$ I- $Eco$ RI fragment shown. The initiator ATG of the  $ChiP$  gene was also mutated to an  $Nco$ I site. The structure of the product was confirmed by DNA sequencing. The 4.9-kb  $Hind$ III- $Eco$ RI construct was cloned into pBIN19 (ref. 24) for plant transformation by either the leaf disk<sup>16</sup> or hypocotyl<sup>27</sup> methods.

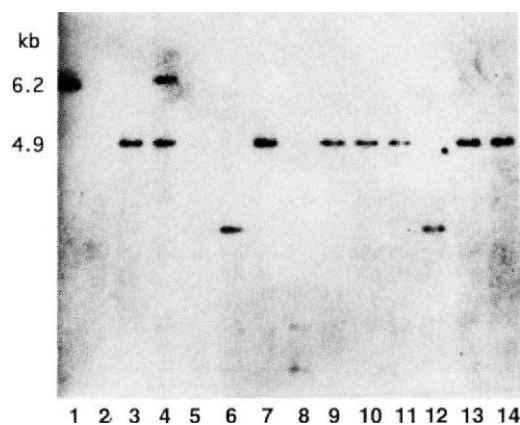


FIG. 2 Detection of the chimaeric  $ChiP$ - $S_{A2}$ -RNase gene in DNA from putative transgenic plants. Sequences to homologous  $S_{A2}$  were detected by hybridization. Lane 1, *N. alata*  $S_{A2}S_{A2}$ ; lane 2, untransformed *N. langsdorffii*  $\times$  *N. alata* hybrid; lanes 3-14, putative transgenic plants. The 6.2-kb band in lane 1 corresponds to the  $Eco$ RI fragment described in Fig. 1 legend. No strongly hybridizing bands were detected in the untransformed hybrid. With the exception of plant 5 (lane 5), each of the transformants contains a restriction fragment that hybridizes to the  $S_{A2}$ -RNase probe. The 4.9-kb band in lanes 3, 4, 7, 9-11, 13 and 14 corresponds to an intact copy of the chimaeric gene. Smaller bands in lanes 6, 8 and 12 suggest that these plants contain truncated copies of the transgene. Detection of an additional restriction fragment in lane 4 indicates that this plant contains at least one additional copy of the chimaeric gene.

**METHODS.** Genomic DNAs were isolated as described<sup>28</sup> and digested with  $Hind$ III and  $Eco$ RI. DNAs (10  $\mu$ g) were separated in a 0.7% agarose gel and blotted onto a nylon membrane. A <sup>32</sup>P-labelled antisense RNA probe derived from the 3' portion of  $S_{A2}$ -RNase cDNA was hybridized to the blot at  $7 \times 10^7$  c.p.m. ml<sup>-1</sup> as described<sup>29</sup>. The blot was autoradiographed at -70 °C for 30 h.

tract<sup>15</sup>, and pollen rejection may be related to action of  $S$ -RNase on pollen RNAs<sup>10</sup>. Efforts to understand the molecular basis for pollen recognition and rejection have been limited by the lack of a system for manipulating and expressing  $S$ -RNases<sup>4-6,16</sup>. Here we use the promoter of a style-expressed gene from tomato to obtain high levels of  $S$ -RNase expression in transgenic *Nicotiana*. Recognition and rejection of *N. alata* pollen  $S$ -alleles occur faithfully in the transgenic plants. Our results show that  $S$ -RNases alone are sufficient for pollen rejection in this system.

A self-incompatible horticultural accession of *N. alata* was the source for two  $S$ -alleles,  $S_{A2}$  and  $S_{C10}$ . The corresponding  $S$ -RNases were analysed biochemically, complementary cDNA clones were obtained, and a genomic clone containing the  $S_{A2}$ -RNase coding region was isolated. The ribonuclease-specific activities and cDNA sequences were similar to those of previously described *N. alata*  $S$ -alleles<sup>9,17,18</sup>. The  $S_{A2}$ -RNase gene

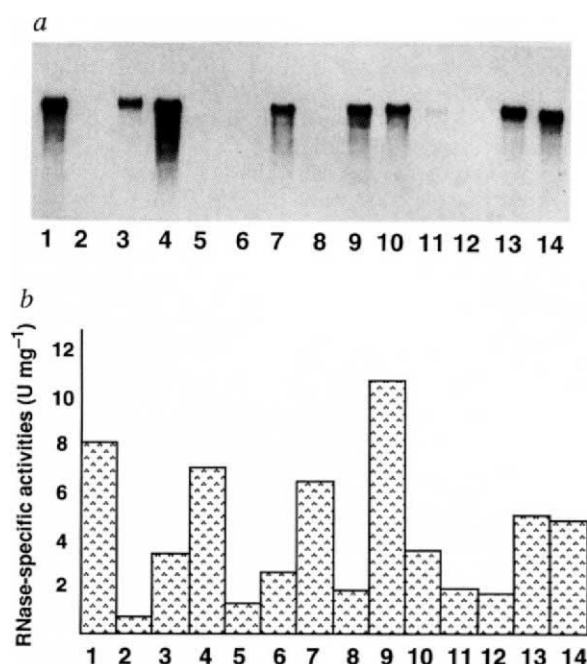


FIG. 3 Analysis of transgene expression. *a*, Detection of  $S_{A2}$ -homologous transcripts on an RNA blot. Lanes are numbered as in Fig. 2. The  $S_{A2}$  transcript is detected in the eight plants containing intact copies of the transgene. *b*, Histogram of RNase activity in style extracts. Plants are numbered as in the other figures. Extracts were prepared and assayed as described<sup>9</sup>. Seven plants (3, 4, 7, 9, 10, 13 and 14) showed high  $S_{A2}$  transcript levels and RNase-specific activities. The specific activity of the extract from *N. alata*  $S_{A2}S_{A2}$  (plant 1; 7.8 U per mg protein) was over tenfold higher than that of the untransformed control (plant 2; 0.75 U per mg protein). *c*, Identity of  $S_{A2}$ -RNase from the transgenics with authentic  $S_{A2}$ -RNase. SDS-PAGE analysis of authentic  $S_{A2}$ -RNase from *N. alata*  $S_{A2}$  homozygote (lane 1; 1.6  $\mu$ g) and  $S_{A2}$ -RNase from pooled transgenic plant styles (lane 2; 0.6  $\mu$ g). The proteins co-migrate with an apparent  $M_r$  of 29K. Similar RNase-specific activities (83 and 86 U per mg protein, respectively) and identity of the N-terminal sequences (DFDYMQLVLT) confirms the identity of the two proteins.

**METHODS.** Total RNAs from mature styles were separated in a 2% agarose-formaldehyde gel (2.5  $\mu$ g per lane), blotted, and probed for  $S_{A2}$ -homologous transcripts as described for Fig. 2. The autoradiograph was exposed at room temperature for 15 min.  $S_{A2}$ -RNase purification was by a method adapted from ref. 14. The N-terminal sequences were analysed using an Applied Biosystems model 470A sequencer.

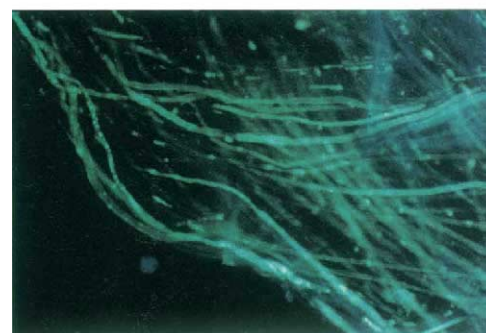
*a*

| Plant number          | 1 | 2   | 3          | 4          | 5   | 6   | 7          | 8   | 9          | 10          | 11  | 12  | 13         | 14         |
|-----------------------|---|-----|------------|------------|-----|-----|------------|-----|------------|-------------|-----|-----|------------|------------|
| $S_{A2}$ pollination  | * | 6/6 | <u>5/8</u> | <u>0/5</u> | 5/5 | 3/3 | <u>2/6</u> | 5/6 | <u>0/8</u> | <u>3/10</u> | 3/3 | 4/4 | <u>2/6</u> | <u>1/7</u> |
| $S_{C10}$ pollination | * | 5/5 | 4/4        | 7/7        | 5/6 | 3/3 | 7/7        | 5/5 | 8/8        | 5/5         | 3/3 | 4/4 | 6/6        | 5/5        |

FIG. 4 Pollination phenotypes of the transgenic plants. *a*, Plants numbered as in Figs 2, 3*a* and *b*, were tested for compatibility with *N. alata*  $S_{A2}$  and  $S_{C10}$  pollen. Data are presented as the number of pollinations that set fruit over total pollinations attempted.  $S_{A2}S_{A2}$  homozygotes in one population show 6/168 fruits set after  $S_{A2}$  pollination and 150/171 fruits set after  $S_{C10}$  pollination (asterisks). Seven of the transgenic plants (underlined) were either partially or fully incompatible with the  $S_{A2}$  pollen. The same 7 plants contained relatively high levels of  $S_{A2}$ -RNase transcript and specific RNase activities in their styles (Fig. 3). All plants set seed when pollinated with  $S_{C10}$  pollen. *b*, Morphology of pollen tubes growing in styles of transgenic plant 9. Style squashes were stained with decolorized aniline blue and viewed by epifluorescence as described<sup>30</sup>. Top, compatible  $S_{C10}$  pollen tubes, characterized by thin walls (green fluorescence) and regularly spaced callose plugs (brightly stained spots), are observed near the base of the style at 42 h post-pollination. Centre, incompatible  $S_{A2}$  pollen tubes, observed ~1 cm below the stigma at 42 h post-pollination, are characterized by thickened, brightly stained walls, infrequent callose plugs, and disorganized growth. Bottom,  $S_{A2}$  pollen tubes at the base of the style at 50 h post-pollination (green fluorescence). Nonspecific staining of epidermal and transmitting tract tissue is also visible in this field. Observations of four similar styles showed 4, 4, 6 and 24  $S_{A2}$  pollen tubes, respectively, penetrating to the base of the style. The morphologies of the compatible and incompatible pollen tubes are typical of the SI phenotype observed in the Solanaceae<sup>4</sup>.



*N. alata*  $S_{C10}$  pollen  
Lower style



*N. alata*  $S_{A2}$  pollen  
Upper style



*N. alata*  $S_{A2}$  pollen  
Lower style

structure was similar to that of other *S*-RNase genes<sup>19,20</sup>. To demonstrate genetic linkage between our cloned sequences and the *S*-locus, an  $F_2$  population was derived by force-selfing an  $S_{A2}S_{C10}$  heterozygote. The plants were tested for rejection of  $S_{A2}$  and/or  $S_{C10}$  pollen, and the presence of cloned  $S_{A2}$ - or  $S_{C10}$ -RNase sequences was determined by polymerase chain reaction (PCR) analysis of genomic DNA. Among 309 plants ( $S_{A2}S_{A2}(67):S_{A2}S_{C10}(165):S_{C10}S_{C10}(77)$ ; 1:2:1  $\chi^2=2.07$ ,  $0.5 \geq P \geq 0.3$ ), pollination phenotype always correlated with the presence of the cloned sequences (that is, no recombination was observed). At the 95% confidence level, these data are consistent with linkage of  $\leq 0.49$  cM between the cloned sequences and the *S*-locus<sup>21</sup>.

Two difficulties hampered previous transformation studies: *N. alata* regenerated in tissue culture with low efficiency<sup>4</sup> and intact *S*-RNase gene clones were not expressed in transgenic plants<sup>4,6</sup>. We have therefore developed an efficient transformation and regeneration system using a heterologous promoter active in the transmitting tract.

A tomato chitinase gene, *ChiP*, originally designated 9617, was identified in a screen for sequences expressed predominantly in the pistil<sup>22,23</sup>. The *ChiP* promoter directs high levels of marker gene expression in the mature transmitting tract as well as in pollen (K. Harikrishna and C.S.G., manuscript in preparation). We therefore constructed the chimaeric  $S_{A2}$ -RNase gene shown in Fig. 1. The construct (containing the 1.4-kilobase (kb) *ChiP* promoter, the  $S_{A2}$ -RNase coding region interrupted by a 114-base-pair (bp) intron, and 2.5 kb of downstream sequence) was inserted in the Ti plasmid vector pBIN19 (ref. 24) as a 4.9-kb *EcoRI*-*HindIII* fragment for plant transformation.

As the transgene recipient, we developed a hybrid between a self-compatible *N. alata* accession (Cultivar Breakthrough; Thompson and Morgan, Jackson, New Jersey, USA) and a close

relative, *Nicotiana langsdorffii* (Digger's Seeds, Dromona, Victoria, Australia). This hybrid was superior to either parent for characteristics such as regeneration in tissue culture and pistil size. The hybrids were fertile and uniform in appearance. Genetic studies indicated that the hybrid could be used as a model for function of the style part of the *N. alata* self-incompatibility (SI) system. For example, a single hybrid plant was crossed with *N. alata* genotypes  $S_{A2}S_{A2}$  and  $S_{C10}S_{C10}$ . The progeny (either  $S_{A2}S_0$



or  $S_{C10}S_0$ ) rejected pollen from the appropriate self-incompatible parent. Furthermore, the two progeny genotypes were sibcrossed to generate an  $F_2$  population. In the 123 plants for which data are available, pollination behaviour correlated only with the presence of the appropriate  $S$ -allele (detected by PCR of genomic DNA or SDS PAGE analysis of style extracts; data not shown). Therefore it is unlikely that the hybrid contains genetic factors that segregate and interfere with style  $S$ -gene function.

DNA gel-blot analysis of putative transformants (Fig. 2) showed that, of twelve kanamycin-resistant plants, eight contained intact copies of the 4.9-kb *EcoRI*–*HindIII* *ChiP*  $S_{A2}$ -RNase construct. Further analysis suggested that each of these plants represents a separate transformation event because DNA digested with *EcoRI* alone revealed fragments of unique sizes (data not shown). High levels of  $S_{A2}$  transcripts were detected in mature styles of seven plants (Fig. 3a; plants 3, 4, 7, 9, 10, 13 and 14). The styles of these seven plants also had high RNase-specific activity (Fig. 3b). These seven plants averaged 75% of the stylar RNase-specific activity of a *N. alata*  $S_{A2}S_{A2}$  homozygote.

Mature styles from the transgenic plants were pooled to obtain sufficient material for protein purification. Mono-S FPLC of the pooled extract showed a peak eluting at the same position as authentic  $S_{A2}$ -RNase. The protein purified from the transgenic plants and authentic  $S_{A2}$ -RNase co-migrated in SDS-PAGE (Fig. 3c) and gave RNase-specific activities of 86 and 83 units per mg protein respectively. In addition, the two proteins had identical N-terminal sequences, consistent with the sequence of the *ChiP*– $S_{A2}$ -RNase construct. We conclude that the product in the transgenic plants is identical to authentic  $S_{A2}$ -RNase.

Pollinations were performed to test for  $S$ -allele-specific pollen rejection. All plants were compatible with  $S_{C10}$  pollen, and plants 4, 9, 10 and 13 set seed when pollinated with up to three additional  $S$ -alleles. But the seven plants expressing  $S_{A2}$ -RNase at relatively high levels rejected  $S_{A2}$  pollen in some or all of the trials (Fig. 4a). The two plants (4 and 9) showing highest  $S_{A2}$ -RNase expression rejected  $S_{A2}$  pollen in all of at least five trials. Pollen tubes in the style of plant 9 were examined by fluorescence microscopy (Fig. 4b): *N. alata*  $S_{C10}$ -pollen tubes were healthy throughout the style (Fig. 4b, top); however, *N. alata*  $S_{A2}$ -pollen tubes, photographed in the top part of the style, showed typically incompatible morphology (Fig. 4b, centre); very few incompatible tubes reached the bottom of the style (Fig. 4b, bottom). In control pollinations of untransformed plants, pollen tubes of both genotypes appeared to be healthy (data not shown).

Our results provide direct evidence that the *N. alata*  $S$ -RNase gene is sufficient for allele-specific recognition and rejection of *N. alata* pollen. Using the system we have developed, it should be possible to dissect the functional elements of  $S$ -RNase genes that are involved in cell–cell communication. □

20. Kaufmann, H., Salamini, F. & Thompson, R. D. *Molec. gen. Genet.* **226**, 457–466 (1991).
21. Stevens, W. L. *J. Genet.* **43**, 301–307 (1942).
22. Gasser, C. S., Budelier, K. A., Smith, A. G., Shah, D. M. & Fraley, R. T. *Pl. Cell* **1**, 15–24 (1989).
23. Gasser, C. S., Harikrishna, K., Jampates-Beale, R., Milligan, S. B. & Robinson-Beers, K. in *Molecular Biology of Tomato* (ed. Yoder, J.) 161–173 (Technomic, Lancaster, Pennsylvania, 1993).
24. Bevan, M. *Nucleic Acids Res.* **12**, 8711–8721 (1984).
25. Clark, K. R., Okuley, J. J., Collins, P. D. & Sims, T. L. *Pl. Cell* **2**, 815–826 (1990).
26. Ai, Y. et al. *Sex. Pl. Reprod.* **3**, 130–138 (1990).
27. Ebert, P. R. & Clarke, A. E. *Pl. molec. Biol.* **14**, 815–824 (1990).
28. Bernatzky, R. & Tanksley, S. D. *Theor. appl. Genet.* **72**, 314–321 (1986).
29. McClure, B. A. & Guilfoyle, T. *Science* **243**, 91–93 (1989).
30. Kho, Y. O. & Baer, J. *Euphytica* **17**, 299–302 (1968).

ACKNOWLEDGMENTS. We thank J. T. Walters for technical assistance; S. A. McMains for help with genetic analysis; E. H. Coe, G. Johal and J. D. Wall for discussion; and T. E. Phillips and M. Stanley for help with fluorescence microscopy. This work was supported by the University of Missouri Research Board and the Food for the 21st Century Program.

## Evidence for a common origin of chloroplasts with light-harvesting complexes of different pigmentation

G. R. Wolfe\*, F. X. Cunningham\*, D. Durnford†, B. R. Green† & E. Gantt\*

\* Department of Botany, University of Maryland, College Park, Maryland 20742, USA

† Botany Department, University of British Columbia, Vancouver, British Columbia V6T 1Z4, Canada

THE red algae (Rhodophyta), which like cyanobacteria have only chlorophyll *a* and use phycobilisomes for light-harvesting<sup>1,2</sup>, are often considered to have originated independently of other photosynthetic eukaryotes, namely the chlorophyll *a/b*-containing Chlorophyta and the chlorophyll *a/c*-containing Chromophyta<sup>3</sup>. Here we report that the red alga *Porphyridium cruentum* has a chlorophyll *a*-containing antenna complex functionally associated with photosystem I, and that polypeptides of this antenna complex are immunologically related to those of higher-plant chlorophyll *a/b* complexes and to those of chromophyte fucoxanthin-chlorophyll *a/c* antenna complexes. This establishes a clear link between organisms containing phycobilisomes and those containing chlorophyll-based light-harvesting complexes and shows that these antennae can co-exist in the same organism. Furthermore, it suggests that the light-harvesting proteins of all photosynthetic eukaryotes had a common origin and supports the idea that chloroplasts had a common ancestor<sup>4–6</sup>.

Photosystem I (PS I) holocomplex was isolated from *P. cruentum* (ATCC 50161) thylakoids by mild detergent solubilization with  $\beta$ -dodecyl maltoside after removal of extrinsic membrane proteins<sup>7</sup>. Pigment analysis by HPLC confirmed that chlorophyll *a* was the only chlorophyll type in either the PS I holocomplex (Table 1) or intact thylakoids<sup>8</sup>. Treatment of the holocomplex with a higher concentration of  $\beta$ -dodecyl maltoside, followed by sucrose gradient separation, yielded two minor and two major chlorophyll-containing fractions. According to their characteristics, the major fractions were designated as light-harvesting complex I (LHC I) and PS I core. The LHC I, like that of green plants<sup>9,10</sup>, was highly enriched in xanthophylls (75 zeaxanthin molecules per 100 chlorophyll *a* molecules) and had a prominent absorbance in the carotenoid region (460–520 nm). In contrast, the PS I core complex was enriched in  $\beta$ -carotene (18  $\beta$ -carotene molecules per 100 chlorophyll *a* molecules), a feature of other PS I core complexes<sup>10</sup>. Whereas the PS I holocomplex had 130 chlorophyll/P700, the core complex after removal of the LHC I complex had 100 chlorophyll/P700, similar to PS I core complexes of cyanobacteria and higher plants<sup>9,10</sup>.

The *P. cruentum* LHC I contained several polypeptides of  $M_r$

Received 7 September; accepted 29 December 1993.

1. de Nettancourt, D. *Incompatibility in Angiosperms* (Springer, New York, 1977).
2. Haring, V., Gray, J. E., McClure, B. A., Anderson, M. A. & Clarke, A. E. *Science* **250**, 937–941 (1990).
3. Nasrallah, J. B., Nishio, T. & Nasrallah, M. E. *Rev. Pl. Physiol. Pl. molec. Biol.* **42**, 393–422 (1991).
4. Murfett, J. & McClure, B. A. in *Molecular Biology of Tomato* (ed. Yoder, J.) 139–151 (Technomic, Lancaster, Pennsylvania, 1993).
5. Sims, T. L. *Crit. Rev. Pl. Sci.* **12**, 129–167 (1993).
6. Thompson, R. D. & Kirch, H.H. *Trends Genet.* **8**, 381–387 (1992).
7. Stein, J. C., Howlett, B., Boyes, D. C., Nasrallah, M. E. & Nasrallah, J. B. *Proc. natn. Acad. Sci. U.S.A.* **88**, 8816–8820 (1991).
8. Goring, D. R., Glavin, T. L., Schafer, U. & Rothstein, S. J. *Pl. Cell* **5**, 531–539 (1993).
9. McClure, B. A. et al. *Nature* **342**, 955–957 (1989).
10. McClure, B. A., Gray, J. E., Anderson, M. A. & Clarke, A. E. *Nature* **347**, 757–760 (1990).
11. Brothtaerts, W., Vanvinckenroye, P., Decock, B., Van Damme, J. & Vendrig, J. C. *Sex. Pl. Reprod.* **4**, 258–266 (1991).
12. Ioerger, T. R., Gohlke, J. R., Xu, B. & Kao, T-H. *Sex. Pl. Reprod.* **4**, 81–87 (1991).
13. Singh, A., Ai, Y. & Kao, T-H. *Pl. Physiol.* **96**, 61–68 (1991).
14. Anderson, M. A. et al. *Nature* **31**, 38–44 (1986).
15. Jahnen, W., Batterham, M. P., Clarke, A. E., Moritz, R. L. & Simpson, R. J. *Pl. Cell* **1**, 493–499 (1989).
16. Murfett, J. et al. *Pl. Cell* **4**, 1063–1074 (1992).
17. Kheyr-Pour, A. et al. *Sex. Pl. Reprod.* **3**, 88–97 (1990).
18. Anderson, M. A. et al. *Pl. Cell* **1**, 483–491 (1989).
19. Coleman, C. E. & Kao, T-H. *Pl. molec. Biol.* **18**, 725–737 (1992).

TABLE 1 Characteristics of photosystem I chlorophyll complexes from *Porphyridium cruentum*

|                                    | Holocomplex | LHC I         | PS I core |
|------------------------------------|-------------|---------------|-----------|
| Chl/P700                           | 132 ± 12    | —             | 98 ± 4    |
| Absorbance maximum (nm)            | 680         | 671, 460–520* | 679       |
| Fluorescence emission maximum (nm) | 730, 680†   | 680           | 720, 680† |
| Chl <i>a</i> :zeaxanthin:          |             |               |           |
| β-carotene (mol:mol:mol)           | 100:9:18    | 100:75:12     | 100:5:18  |

PS I holocomplex was isolated from thylakoids solubilized for 30 min at 4 °C in 0.66% β-dodecyl maltoside (β-DM)/50 mM MES/10 mM NaCl/0.7 M sucrose, pH 6.1, by sucrose gradient centrifugation and anion-exchange columns<sup>7</sup>. LHC I and PS I core preparations were generated from holocomplex samples (600 μg Chl per ml) treated for 1 h with 1% β-DM (w/v) and separated on a 5–30% (w/v) linear sucrose gradient containing 50 mM MES (pH 6.1), 10% glycerol (v/v) and 0.04% β-DM (run for 16 h at 154,000g). Fluorescence emission spectra, with excitation at 440 nm, were recorded at 77 K. Molar pigment analysis by HPLC and Chl/P700 determinations were made as in ref. 8. The Chl/P700 values are means ± s.d.

\* Shoulder; † minor peak.

18–23.5K (bracketed in Fig. 1), which is in the molecular weight range of higher-plant LHC I (ref. 10) and diatom fucoxanthin-chlorophyll *a/c*-binding polypeptides<sup>11</sup>. Its absorption and fluorescence emission maxima were similar to those of higher plants<sup>10</sup>. Given the absence of PS I reaction-centre polypeptides (broad band corresponding to *M<sub>r</sub>* 57–62K in Fig. 1, lane 2), we conclude that chlorophyll *a* is bound by one or more of the 18–23.5K polypeptides. Removal of these polypeptides from the PS I holocomplex correlated with a shift of the fluorescence emission maximum at 77K from 730 nm in the holocomplex to 720 nm in the core complex; the isolated LHC I had a strong emission at 680 nm. These findings indicate that functional coupling and energy transfer between the LHC I complex and the PS I core occurs in the holocomplex. The minor emission at 680 nm in the PS I core complex probably reflects small amounts of LHC I still remaining with the core (Fig. 1, lane 3). The *P. cruentum* LHC I apoproteins give prominent bands on silver-

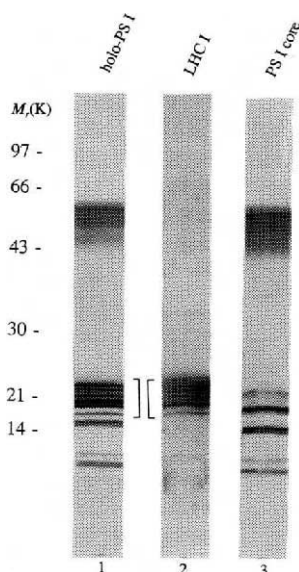


FIG. 1 Isolated photosystem I holocomplex from *Porphyridium cruentum* before and after fractionation (for methods, see Table 1). Chlorophyll (Chl) complexes were treated with SDS and resolved on a 7–15% polyacrylamide gradient gel by SDS-PAGE. Lanes: 1, PS I holocomplex (1.5 μg Chl); 2, LHC I complex (0.5 μg Chl); 3, PS I core (1.5 μg Chl). The LHC I polypeptide region is bracketed. Samples were stained with silver<sup>25</sup>. Molecular size markers are indicated on the left.

stained gels of the PS I holocomplex (Fig. 1, lane 1), indicating that they constitute a significant proportion of the PS I polypeptides.

The relatedness of LHC I polypeptides of *P. cruentum* to those of other photosynthetic eukaryotes was established by immunoblot analysis (Fig. 2). At least five red algal PS I polypeptides of *M<sub>r</sub>*s 19.5, 20, 22, 23 and 23.5K (Fig. 2, lane 2; arrowheads) were immunodecorated with an antibody raised against a barley PS I complex (CP Ia), composed of PS I reaction centre (CP I) plus LHC I (ref. 12). Two additional crossreacting polypeptides of 19.0 and 20.5K were present in whole thylakoids of *P. cruentum* (Fig. 2, lane 1). In a crude spinach PS I preparation, polypeptides of LHC I as well as CP 29 and LHC II were immunodecorated (Fig. 2, lane 3), because this particular antiserum crossreacts with all of the chlorophyll *a/b*-binding polypeptides as a result of shared epitopes<sup>12</sup>. Significantly, the antiserum did not detect polypeptides within the 18–23.5K region in thylakoids of the cyanobacterium *Nostoc* sp. (Fig. 2, lane 4) which were used as a control. It did detect the PS I reaction centre apoproteins (broad band at *M<sub>r</sub>* 60–70K) in *Nostoc* and in all other samples, as expected. The lack of a light-harvesting complex reaction in *Nostoc* confirms the reported absence of these polypeptides in other cyanobacteria<sup>2</sup>, including thermophiles and mesophiles such as *Synechocystis* PCC6803 (refs 13, 14), *Mastigocladus laminosus*<sup>15</sup> and *Synechococcus* PCC6301 (ref. 16).

LHC polypeptides have not previously been found in red algal thylakoids<sup>2,17,19</sup>, probably because the methods used for PS I isolation resulted in considerable loss of chlorophyll-binding polypeptides and, by analogy with cyanobacteria, they were not

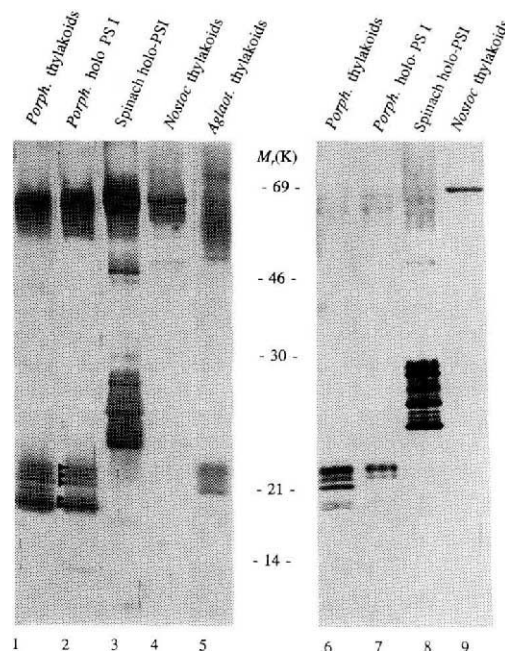


FIG. 2 Immunoblot analysis of photosynthetic membrane components of eukaryotes and a prokaryote. Samples were dissociated with SDS, loaded at 0.5 μg Chl per lane and resolved on SDS-PAGE (12% acrylamide). After electrotransfer to Immobilon-P they were immunostained with an antiserum to barley CP Ia<sup>12</sup> (left panels) or with an antiserum to fucoxanthin-Chl *a/c* complex from a diatom (*P. tricornutum*)<sup>11</sup> (right panel). Arrowheads indicate LHC I polypeptides of the *P. cruentum* PS I holocomplex. The sample source is indicated at the top of each lane: 1, *P. cruentum* thylakoids; 2, PS I holocomplex of *P. cruentum*; 3, PS I holocomplex of spinach; 4, thylakoids of *Nostoc* sp.; 5, thylakoids of *A. neglectum*. Lanes 6–9 were loaded with samples identical to those in lanes 1–4. Molecular size standards are indicated in the centre.



expected. The existence of LHC I polypeptides is not unique to *P. cruentum* because we found a similar group of polypeptides that crossreacted with antibody against CP Ia in thylakoids of another red alga, *Aglaothamnion neglectum* (Fig. 2, lane 5). These two species span the taxonomic spectrum in that *P. cruentum* is a member of the primitive subclass Bangiophyceae (order Porphyridiales), whereas *Aglaothamnion* (formerly *Callithamnion*) belongs to an advanced subclass Floridophycidae (order Ceramiales)<sup>20</sup>.

In addition to being related to higher-plant LHC I, the *P. cruentum* polypeptides are also related to chromophyte chlorophyll *a/c* LHCs. In *P. cruentum* thylakoids, five polypeptides were immunodecorated with an antiserum raised against a fucoxanthin-chlorophyll *a/c*-binding complex purified from a diatom (*Phaeodactylum tricornutum*)<sup>11</sup> (Fig. 2, lane 6). These five are comparable in size to those that reacted with the barley antibody. In the PS I holocomplex at least three comparable polypeptides reacted with the diatom antibody (Fig. 2, lane 7). Interestingly, the 20K and 23.5K polypeptides that reacted with anti-CP Ia were not detected in *P. cruentum* with the diatom antibody. Spinach LHC I polypeptides (23–26K), LHC II and CP29 polypeptides all reacted strongly with this antibody (Fig. 2, lane 8), confirming the relatedness of chlorophyll *a/b*- and *a/c*-binding polypeptides<sup>21</sup>. With this antibody there was no crossreactivity with *Nostoc* polypeptides, except for a faint band of unknown identity at  $M_r \sim 65$ K. We also tested the prochlorophyte *Prochlorothrix hollandica* with the barley anti-CP Ia antiserum and did not find any crossreactivity with LHC-type polypeptides (Fig. 3), confirming the known lack of immunological relatedness between the chlorophyll *a/b* polypeptides of this group of prokaryotes<sup>22</sup> and the chlorophyll *a/b* polypeptides of eukaryote chloroplasts.

The immunological relatedness of LHCs from groups as phylogenetically distinct as rhodophytes, chlorophytes and chromophytes indicates conservation of primary structure during evolution. It strongly suggests that the light-harvesting proteins of all photosynthetic eukaryotes had a common ancestor and supports the concept of a common evolutionary origin for all chloroplasts. To our knowledge, this is the first demonstration of an LHC-type antenna system in phycobilisome-containing organisms and clearly establishes that two very different types of antenna systems exist in a single chloroplast. Our findings lend credence to the hypothesis that an ancestral chloroplast may have contained both phycobilisomes and LHCs<sup>23</sup>. Subsequent loss of one or the other antenna system by various groups of organisms could explain the chloroplast diversity that exists today. Our results support the concept of a monophyletic origin of the chloroplast<sup>5,6</sup> rather than the polyphyletic origin<sup>3</sup> initially proposed on the basis of widely different types of light-harvesting antennas. This view has recently received support from the reported similarities between chloroplast genes of the rhodophyte *Porphyra*<sup>24</sup> and those of chromophyte and chlorophyte lineages<sup>6</sup>, although genes for LHCs have so far not been found in chloroplast genomes. □

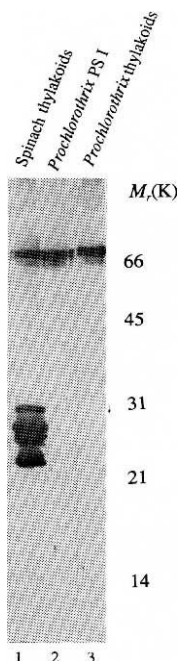


FIG. 3 Immunoblot probed with antiserum to barley CP Ia<sup>12</sup> (methods as for Fig. 2). The sample source is indicated at the top of each lane: lane 1, spinach thylakoids; lane 2, *P. hollandica* PS I-enriched fraction<sup>26</sup>; and lane 3, thylakoids of *P. hollandica*.

Received 9 June; accepted 19 November 1993.

- Gantt, E. *Int. Rev. Cytol.* **66**, 45–80 (1980).
- Glazer, A. N. & Melis, A. A. *Rev. Pl. Physiol.* **38**, 11–45 (1987).
- Raven, P. H. *Science* **189**, 641–646 (1970).
- Taylor, F. J. R. *Proc. R. Soc. Lond. B* **204**, 267–286 (1979).
- Cavalier-Smith, T. *BioSystems* **28**, 91–106 (1992).
- Morden, C. W., Delwiche, C. F., Kuhse, M. & Palmer, J. D. *BioSystems* **28**, 75–90 (1992).
- Wolfe, G. R., Cunningham, F. X. & Gantt, E. *Research in Photosynthesis 1* **Proc. IX Int. Cong. Photosynthesis**, Nagoya, Japan (ed. Murata, N.) 315–318 (Kluwer, Dordrecht, 1990).
- Cunningham, F. X., Dennenberg, R. J., Mustardy, L., Jursinic, P. A. & Gantt, E. *Pl. Physiol.* **91**, 1179–1187 (1989).
- Bassi, R., Rigoni, F. & Giacometti, G. M. *Photochem. Photobiol.* **52**, 1187–1206 (1990).
- Thornber, J. P., Morishige, D. T., Anandan, S. & Peter, G. F. in *Chlorophylls* (ed. Scheer, H.) 549–585 (CRC Press, Boca Raton, 1991).
- Fawley, M. W. & Grossman, A. R. *Pl. Physiol.* **81**, 149–155 (1986).
- White, M. J. & Green, B. R. *Eur. J. Biochem.* **163**, 545–551 (1987).
- Krauss, N. et al. *Nature* **361**, 326–331 (1993).
- Roegner, M., Nixon, P. J. & Diner, B. A. *J. Biol. Chem.* **265**, 6189–6196 (1990).
- Almog, O., Shoham, G., Michaeli, D. & Nechustai, R. *Proc. natn. Acad. Sci. U.S.A.* **88**, 5312–5320 (1991).
- Lundell, D. J., Glazer, A. N., Melis, A. & Malkin, R. J. *Biol. Chem.* **260**, 646–654 (1985).
- Redlinger, T. & Gantt, E. *Pl. Physiol.* **73**, 36–40 (1983).
- Marquardt, J. & Ried, A. *Planta* **187**, 372–380 (1992).
- Yurina, N. P., Karakashev, G. V., Karapetyan, N. V. & Odintsova, M. S. *Photosyn. Res.* **30**, 15–23 (1991).
- Lee, R. E. *Phycology* (Cambridge, London, 1980).
- Green, B., Pichersky, E. & Kloppstech, K. *Trends biochem. Sci.* **16**, 181–186 (1991).
- Hiller, R. G. & Larkum, A. W. D. *Biochem. biophys. Acta* **806**, 107–115 (1985).
- Bryant, D. A. *Curr. Biol.* **2**, 240–242 (1992).
- Reith, M. & Munholland, J. *Pl. Cell* **5**, 465–475 (1993).
- Oakley, B. R., Kirsch, D. R. & Morris, N. R. *Analyt. Biochem.* **105**, 361–363 (1980).
- van der Staay, G. W. M., Brouwer, A., Baard, R. L., van Mourik, F. & Matthijs, H. C. P. *Biochem. biophys. Acta* **1102**, 220–228 (1992).

ACKNOWLEDGEMENTS. We thank A. R. Grossman for the gift of diatom LHC antibodies and G. W. M. van der Staay for the *Prochlorothrix* samples. Our work was supported by the US Department of Energy and the Maryland Agricultural Experiment Station (E.G.), and the Natural Sciences and Engineering Council of Canada (B.R.G.).

## Regulation of transcription by dimerization of erythroid factor NF-E2 p45 with small Maf proteins

Kazuhiko Igarashi\*, Kohsuke Kataoka†, Ken Itoh\*, Norio Hayashi\*, Makoto Nishizawa† & Masayuki Yamamoto\*‡

\* Department of Biochemistry, Tohoku University School of Medicine, 2-1 Seiryomachi, Aoba-ku, Sendai 980, Japan

† Department of Viral Oncology, Cancer Institute, 1-37-1 Kami-Ikebukuro, Toshima-ku, Tokyo 170, Japan

TRANSCRIPTION factor NF-E2 is crucial for regulating erythroid-specific gene expression<sup>1</sup>. Cloning of the NF-E2 p45 protein has revealed that it contains a basic region-leucine zipper (b-zip) domain which associates with another unidentified protein (of relative molecular mass 18,000) to form functional NF-E2 (ref. 2). We show here that products of the *maf* proto-oncogene family<sup>3–5</sup>, MafF, MafG and MafK (the small Maf proteins) which possess a b-zip DNA-binding domain but lack a canonical transactivation domain<sup>3</sup>, directly control the DNA-binding properties of p45 by heterodimeric association with p45. Whereas homodimers of the small Maf proteins act as negative regulators, heterodimers composed of Maf and p45 support active transcription *in vivo*. These results indicate that one (or all) of the small Maf proteins is the second constituent chain required for NF-E2 activity, and that negative as well as positive regulation can be achieved through an NF-E2 site, depending on the equilibrium concentrations of p45 and the Maf proteins inside erythroid cells.

Through binding-site selection analysis, the consensus binding site for v-Maf was found to be an unusually long 13-base-

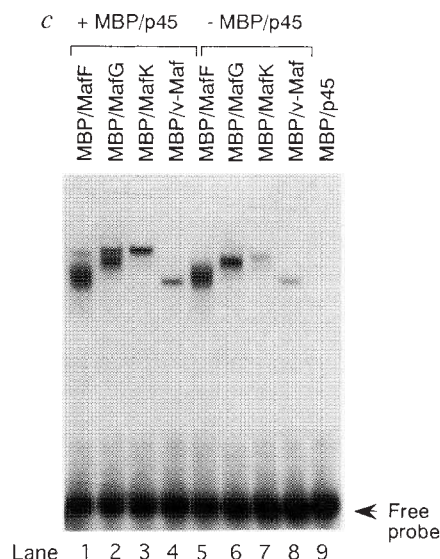
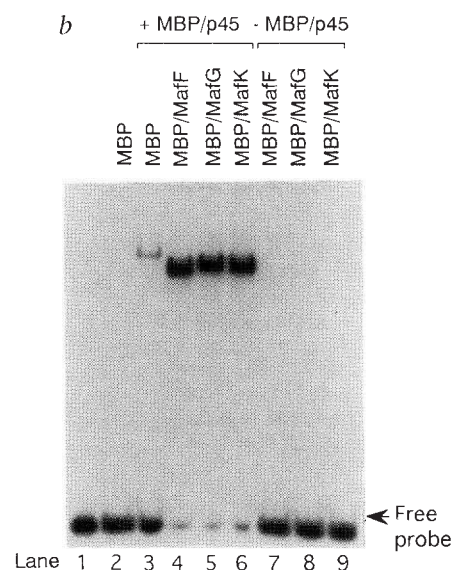
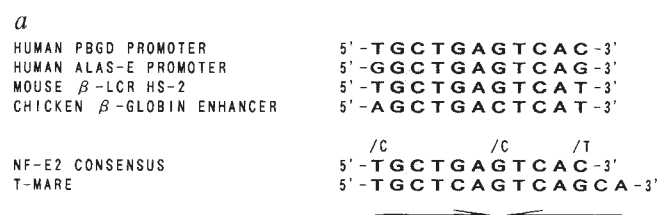
‡ To whom correspondence should be addressed.

pair (bp) (TRE (12-*O*-tetradecanoate-13-acetate-responsive element)-type *Maf* recognition element, T-MARE or 14-bp palindromic sequence which is also bound by proteins encoded by other members of the *maf* gene family<sup>6</sup>. The T-MARE is very similar to the consensus for NF-E2 (ref. 2) (Fig. 1a). Because of this binding-site sequence similarity, we investigated whether a *Maf* protein could be the heterodimeric partner of p45. *Maf* proteins and p45 were expressed in *Escherichia coli* as fusions with maltose-binding protein (MBP), and binding to NF-E2 sites was examined using electrophoretic gel mobility shift assay. Incubation of a probe containing the NF-E2 binding site within the human porphobilinogen deaminase promoter (PBGD probe<sup>7</sup>) with purified p45 resulted in a retarded band of low intensity (Fig. 1b, lane 3). We detected only marginal binding to the PBGD probe using bacterially expressed MafF, MafG or MafK (Fig. 1b, lanes 7–9). In contrast, addition of any of the small *Maf* proteins to p45 generated much more intense signals from nucleoprotein complexes with different mobilities, and led to the coincident disappearance of the more slowly migrating p45 homodimer complex (Fig. 1b, lanes 4–6). Formation of the complexes was efficiently prevented by competition with excess unlabelled PBGD probe but not by an unrelated probe, demonstrating that DNA binding was specific (data not shown).

We concluded that these nucleoprotein complexes contained heterodimers of p45 with one or another of the *Maf* family proteins on the basis of the following observations. First, MBP

similarly purified did not affect the binding properties of p45 (Fig. 1b, lane 3) or of the small *Maf* fusion proteins (not shown), excluding the possibility that the MBP part of these fusion proteins was increasing the affinity of p45 or *Maf* for the NF-E2 site. We consistently observed a similar increase in binding activity using *in vitro*-translated p45 and MafF proteins (not shown). Second, the mobility of the complexes formed in the presence of p45 together with any one of MafF, MafG or MafK was greater than that of the complex formed by homodimeric p45 alone (Fig. 1b). As the *Maf* fusion proteins were smaller than the p45 fusion protein, the differences in mobility were probably a result of heterodimer formation. Third, when a mixture of p45 fused to glutathione *S*-transferase (GST-p45) and MBP-MafF was incubated with the PBGD probe, the mobility of the resulting complex was different from that of the GST-p45 homodimer complex. Addition of anti-MBP antiserum prevented the formation of this complex (result not shown), indicating that MBP-MafF was present in the complex. Fourth, when the T-MARE probe was used, complexes were seen whose mobility was distinct from that of the respective *Maf* homodimer complex (Fig. 1c); p45 alone showed no binding to this probe (lane 9). These observations indicate that heterodimers of p45 with the small *Maf* proteins are binding to the PBGD or T-MARE probes.

To compare the relative affinities of the p45 homodimer and p45/*Maf* heterodimers for the NF-E2 site, various amounts of p45 were incubated with the PBGD probe in the presence or



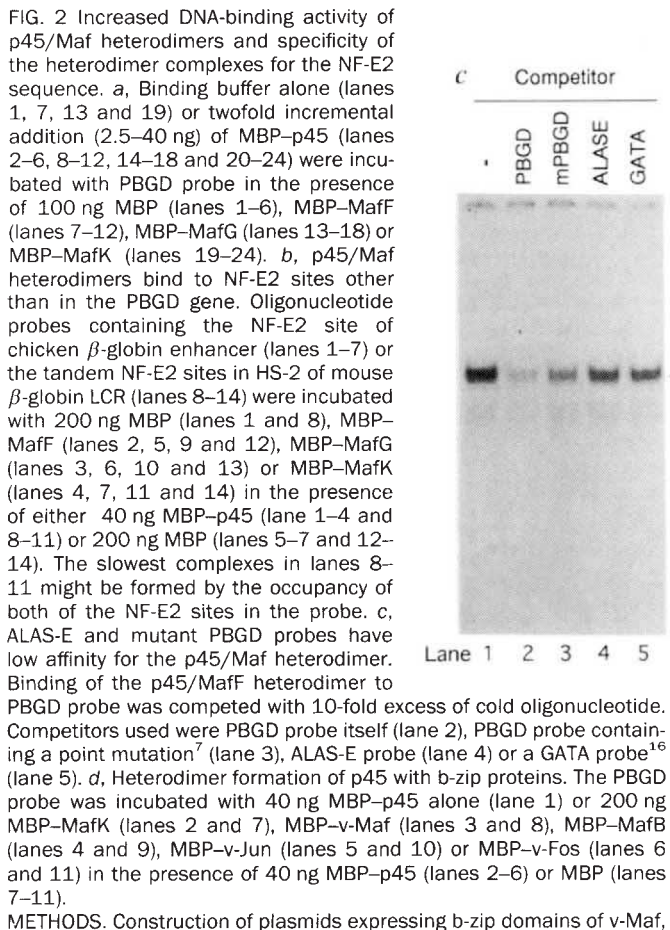
**FIG. 1** Binding of p45/*Maf* heterodimers to NF-E2 site of PBGD gene and T-MARE. **a**, Comparison of NF-E2-binding sequences of human porphobilinogen deaminase (PBGD)<sup>7</sup>, human  $\delta$ -aminolaevulinate synthase (ALAS-E)<sup>10</sup>, DNase I-hypersensitive site 2 (HS-2) of mouse  $\beta$ -globin gene locus control region (LCR)<sup>9</sup> and chicken  $\beta$ -globin gene enhancer<sup>8</sup> with T-MARE. The consensus sequence for NF-E2 binding is from ref. 2. Nucleotides matching the NF-E2 consensus sequence are shaded. Arrows indicate palindromic in T-MARE. **b**, A radiolabelled oligonucleotide containing the NF-E2 site of PBGD (PBGD probe) was incubated with: reaction buffer alone (lane 1), 130 ng MBP (lane 2), 130 ng MBP and 80 ng MBP-p45 (lane 3), 150 ng MBP-MafF (lanes 4 and 7), MBP-MafG (lanes 5 and 8) or MBP-MafK (lanes 6 and 9), in the presence (lanes 4–6) or absence (lanes 7–9) of 80 ng MBP-p45. **c**, A radiolabelled oligonucleotide containing T-MARE (T-MARE probe) was incubated with 40 ng MBP-MafF (lanes 1 and 5), MBP-MafG (lanes 2 and 6), MBP-MafK (lanes 3 and 7) or MBP-v-Maf (lanes 4 and 8) in the presence (lanes 1–4) or absence (lanes 5–8) of 20 ng MBP-p45. The probe was incubated with 20 ng MBP-p45 alone for lane 9.

**METHODS.** A cDNA encoding p45 was generated by polymerase chain reaction (PCR) using cDNAs reverse-transcribed from MEL-cell RNA and a set of primers (5'-ACAGTACCATGGCCCCGTGCTCCTC-3' and 5'-AGACC-AGCTCAATCTGTAGCCTCC-3') designed so that the cDNA has a *Nco*I site (underlined) overlapping the first Met codon of the published sequence<sup>2</sup>. As a result, the second codon (Pro) was substituted with that for Ala. cDNA was cloned into a pBluescript plasmid, and verified by sequencing. The entire open reading frame was cloned in pMALc-2 (New England BioLabs) using the *Nco*I site of the cDNA and the *Eco*RI site of the vector filled in with T4 DNA polymerase. Similarly, cDNAs encoding most of MafF, MafG and MafK were cloned in pMALc (but a small deletion was introduced near the N terminus to give stable expression of the proteins in *E. coli*; M.N. and K.K., unpublished results). These expression plasmids were introduced into *E. coli* strains TB-1 or SG12036, and fusion proteins were purified by amylose resin affinity-column chromatography as recommended by the supplier (New England BioLabs). Electrophoretic gel mobility shift assays were done as described<sup>6</sup> on the MBP fusion proteins using either the PBGD probe (5'-TGGGGAACCTGTGCTGAGTCACTGGAG-3') or the T-MARE probe (5'-GAGTAATGCTGAGTCAGCAATTCGAGCT-3').



An increase in DNA-binding affinity of p45 through heterodimer formation with the small Maf proteins was also observed when we examined other putative NF-E2 sites in the chicken  $\beta$ -globin enhancer<sup>8</sup> and in the locus control region (LCR) of the

|    | <i>b</i> | Chicken | Mouse  |
|----|----------|---------|--------|
| 1  | 0.0000   | 0.0000  | 0.0000 |
| 2  | 0.0000   | 0.0000  | 0.0000 |
| 3  | 0.0000   | 0.0000  | 0.0000 |
| 4  | 0.0000   | 0.0000  | 0.0000 |
| 5  | 0.0000   | 0.0000  | 0.0000 |
| 6  | 0.0000   | 0.0000  | 0.0000 |
| 7  | 0.0000   | 0.0000  | 0.0000 |
| 8  | 0.0000   | 0.0000  | 0.0000 |
| 9  | 0.0000   | 0.0000  | 0.0000 |
| 10 | 0.0000   | 0.0000  | 0.0000 |
| 11 | 0.0000   | 0.0000  | 0.0000 |
| 12 | 0.0000   | 0.0000  | 0.0000 |
| 13 | 0.0000   | 0.0000  | 0.0000 |
| 14 | 0.0000   | 0.0000  | 0.0000 |
| 15 | 0.0000   | 0.0000  | 0.0000 |
| 16 | 0.0000   | 0.0000  | 0.0000 |
| 17 | 0.0000   | 0.0000  | 0.0000 |
| 18 | 0.0000   | 0.0000  | 0.0000 |
| 19 | 0.0000   | 0.0000  | 0.0000 |
| 20 | 0.0000   | 0.0000  | 0.0000 |
| 21 | 0.0000   | 0.0000  | 0.0000 |
| 22 | 0.0000   | 0.0000  | 0.0000 |
| 23 | 0.0000   | 0.0000  | 0.0000 |
| 24 | 0.0000   | 0.0000  | 0.0000 |
| 25 | 0.0000   | 0.0000  | 0.0000 |
| 26 | 0.0000   | 0.0000  | 0.0000 |
| 27 | 0.0000   | 0.0000  | 0.0000 |
| 28 | 0.0000   | 0.0000  | 0.0000 |
| 29 | 0.0000   | 0.0000  | 0.0000 |
| 30 | 0.0000   | 0.0000  | 0.0000 |
| 31 | 0.0000   | 0.0000  | 0.0000 |
| 32 | 0.0000   | 0.0000  | 0.0000 |
| 33 | 0.0000   | 0.0000  | 0.0000 |
| 34 | 0.0000   | 0.0000  | 0.0000 |
| 35 | 0.0000   | 0.0000  | 0.0000 |
| 36 | 0.0000   | 0.0000  | 0.0000 |
| 37 | 0.0000   | 0.0000  | 0.0000 |
| 38 | 0.0000   | 0.0000  | 0.0000 |
| 39 | 0.0000   | 0.0000  | 0.0000 |
| 40 | 0.0000   | 0.0000  | 0.0000 |
| 41 | 0.0000   | 0.0000  | 0.0000 |
| 42 | 0.0000   | 0.0000  | 0.0000 |
| 43 | 0.0000   | 0.0000  | 0.0000 |
| 44 | 0.0000   | 0.0000  | 0.0000 |
| 45 | 0.0000   | 0.0000  | 0.0000 |
| 46 | 0.0000   | 0.0000  | 0.0000 |
| 47 | 0.0000   | 0.0000  | 0.0000 |
| 48 | 0.0000   | 0.0000  | 0.0000 |
| 49 | 0.0000   | 0.0000  | 0.0000 |
| 50 | 0.0000   | 0.0000  | 0.0000 |
| 51 | 0.0000   | 0.0000  | 0.0000 |
| 52 | 0.0000   | 0.0000  | 0.0000 |
| 53 | 0.0000   | 0.0000  | 0.0000 |
| 54 | 0.0000   | 0.0000  | 0.0000 |
| 55 | 0.0000   | 0.0000  | 0.0000 |
| 56 | 0.0000   | 0.0000  | 0.0000 |
| 57 | 0.0000   | 0.0000  | 0.0000 |
| 58 | 0.0000   | 0.0000  | 0.0000 |
| 59 | 0.0000   | 0.0000  | 0.0000 |
| 60 | 0.0000   | 0.0000  | 0.0000 |
| 61 | 0.0000   | 0.0000  | 0.0000 |
| 62 | 0.0000   | 0.0000  | 0.0000 |
| 63 | 0.0000   | 0.0000  | 0.0000 |
| 64 | 0.0000   | 0.0000  | 0.0000 |
| 65 | 0.0000   | 0.0000  | 0.0000 |
| 66 | 0.0000   | 0.0000  | 0.0000 |
| 67 | 0.0000   | 0.0000  | 0.0000 |
| 68 | 0.0000   | 0.0000  | 0.0000 |
| 69 | 0.0000   | 0.0000  | 0.0000 |
| 70 | 0.0000   | 0.0000  | 0.0000 |
| 71 | 0.0000   | 0.0000  | 0.0000 |
| 72 | 0.0000   | 0.0000  | 0.0000 |
| 73 | 0.0000   | 0.0000  | 0.0000 |
| 74 | 0.0000   | 0.0000  | 0.0000 |
| 75 | 0.0000   | 0.0000  | 0.0000 |
| 76 | 0.0000   | 0.0000  | 0.0000 |
| 77 | 0.0000   | 0.0000  | 0.0000 |
| 78 | 0.0000   | 0.0000  | 0.0000 |
| 79 | 0.0000   | 0.0000  | 0.0000 |
| 80 | 0.0000   |         |        |



**d** MBP/p45

| +      |          |           |          |           |           | -        |           |          |           |           |
|--------|----------|-----------|----------|-----------|-----------|----------|-----------|----------|-----------|-----------|
| MBP    | MBP/MafK | MBP/v-Maf | MBP/MafB | MBP/v-Jun | MBP/v-Fos | MBP/MafK | MBP/v-Maf | MBP/MafB | MBP/v-Jun | MBP/v-Fos |
|        |          |           |          |           |           |          |           |          |           |           |
| Lane 1 | Lane 2   | Lane 3    | Lane 4   | Lane 5    | Lane 6    | Lane 7   | Lane 8    | Lane 9   | Lane 10   | Lane 11   |

p45 homodimer →

MBP/p45 + MBP/MafF

← Free probe

NATURE · VOL 367 · 10 FEBRUARY 1994

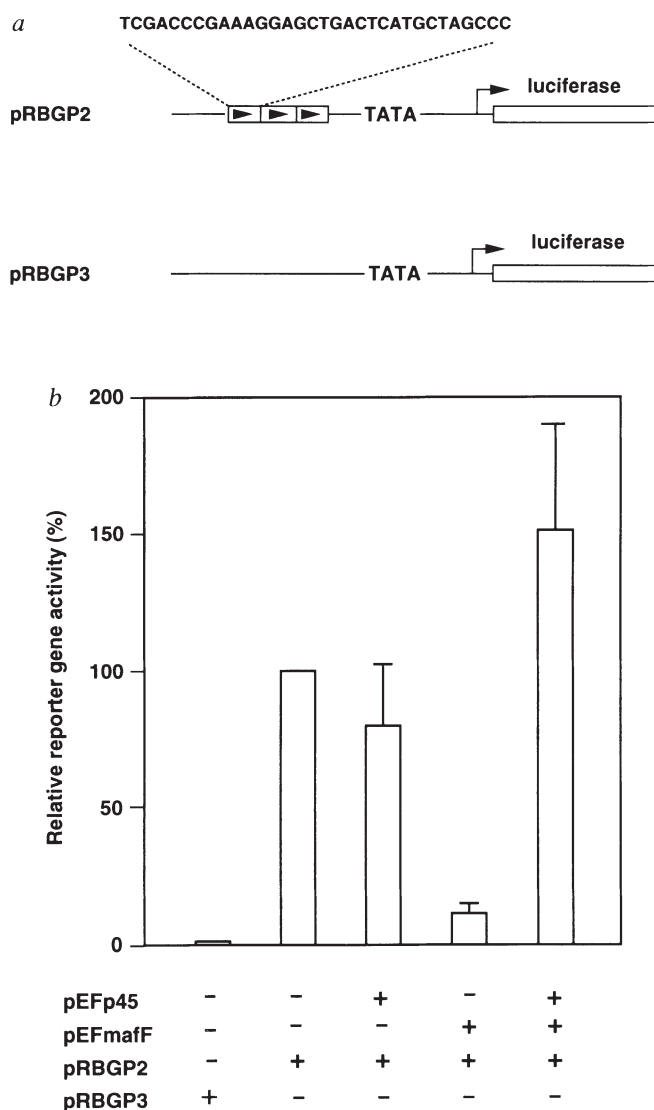


FIG. 3 Activity of the NF-E2 p45 and MafF proteins in NIH3T3 cells. **a**, Structure of the reporter plasmids. Control plasmid pRBGP3 contains the TATA box of the rabbit  $\beta$ -globin gene, whereas plasmid pRBGP2 contains a triplicate NF-E2 binding site of the chicken  $\beta$ -globin enhancer<sup>8</sup>. **b**, Response of reporter gene expression to co-transfection with p45 and/or MafF. The results of four independent experiments are shown, each carried out in duplicate. Luciferase activity of pRBGP2 in the absence of any effector plasmid was set as 100% and relative activities obtained by addition of various combinations of effector plasmids are shown. Standard errors are indicated. The effects of the MafF and p45 expression plasmids were dependent on the NF-E2 sites of the reporter plasmid, because the basal reporter gene activity of pRBGP4, which contains triplicate mutant NF-E2 sites, is only 3% of that of pRBGP2.

**METHODS.** pRBGP3 was constructed by inserting an oligonucleotide carrying the TATA box of rabbit  $\beta$ -globin gene<sup>17</sup> between the *Bgl*II and *Hind*III sites of pGL (Promega). An oligonucleotide containing the chicken  $\beta$ -globin enhancer NF-E2 site or its mutant oligonucleotide (5'-TCGACCCGAAAGGATCTTACTTATGCTAGCCC-3'; mutations underlined) was triplicated and inserted adjacent to the TATA box on pRBGP3 to generate pRBGP2 or pRBGP4, respectively. cDNAs encoding mouse NF-E2 p45 or chicken MafF were inserted into pEF-BssHII, a modified version of the expression vector pEF-BOS<sup>18</sup>, resulting in pEFp45 and pEFmafF, respectively. Transfection was carried out as described<sup>19</sup> using calcium phosphate precipitation with 1  $\mu$ g reporter plasmid and 0.5  $\mu$ g pENL (which expresses  $\beta$ -galactosidase), with or without pEFp45 (1  $\mu$ g) and pEFmafF (0.5  $\mu$ g). Cytoplasmic extracts were prepared 36 h after transfection, and luciferase activities assayed. Relative transfection efficiencies were determined by  $\beta$ -galactosidase assay and used to normalize luciferase activity.

probe to p45/Maf heterodimers, although the affinity of the mutant PBGD probe for p45/Maf is still greater than that of the ALAS-E probe (Fig. 2c, and data not shown). These results are in good agreement with an earlier report that purified NF-E2 does not bind appreciably to ALAS-E or mutant PBGD probes<sup>2</sup>.

The specificity of interaction of proteins containing a b-zip structure with other proteins is usually determined by the structure of the leucine zipper itself<sup>11-13</sup>. Therefore, p45 could interact with v-Maf or MafB, both of which contain putative transactivation domains; however, neither of these proteins formed heterodimers with p45 that could bind DNA (Fig. 2d). It is unlikely that these proteins first form heterodimers which are then unable to bind the probe sequence, because the amount of p45 homodimer complex did not change in the presence of v-Maf or MafB fusion proteins. Similarly, including v-Jun or v-Fos fusion proteins in the reactions did not alter the binding of p45 homodimer (Fig. 2d). In contrast, formation of a p45 homodimer complex was prevented by competition with MafK fusion protein, which can heterodimerize with p45 (as became evident upon longer electrophoresis; Fig. 1b). This result shows that heterodimer formation between p45 and MafF, MafG or MafK is specific, which is consistent with the fact that the primary structures of the b-zip domains of the small Maf proteins are much more similar to one another than they are to either v-Maf or MafB<sup>3</sup>.

The function of p45 and Maf family proteins as transcriptional regulators was assayed by transfecting NIH3T3 cells with luciferase reporter plasmids and with pEFp45 and pEFmafF plasmids expressing p45 and MafF, respectively (Fig. 3). Compared with

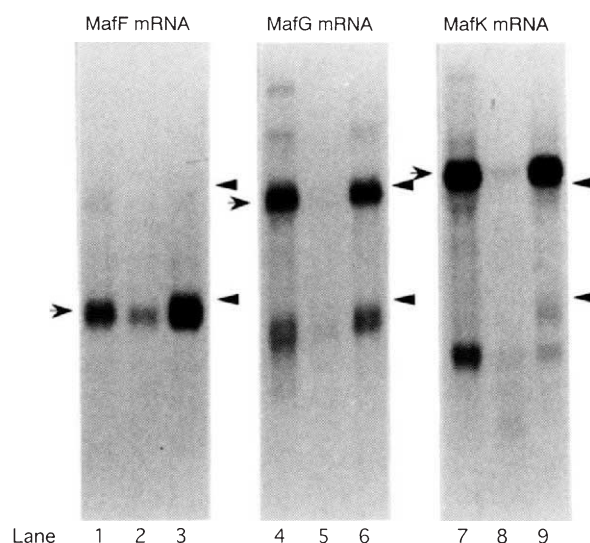


FIG. 4 Expression of MafF, MafG and MafK mRNAs in erythroid and lymphoid cell lines. Poly(A)<sup>+</sup> RNA isolated from HD3 (lanes 1, 4 and 7), QT6 (lanes 2, 5 and 8) and MSB-1 (lanes 3, 6 and 9) cells were separated on a denaturing agarose gel, transferred to nylon membranes and hybridized with MafF, MafG or MafK probes. Arrows and arrowheads indicate signals and the migration positions of ribosomal RNA markers, respectively.

**METHODS.** HD3 is a chicken erythroid cell line transformed with temperature-sensitive avian erythroblastosis virus<sup>14</sup>. MSB-1 is a chicken T-lymphoid cell line<sup>15</sup>, and QT6 is a quail fibroblast cell line<sup>20</sup>. Total RNA was isolated from cultured cells using the guanidine-acidified phenol method<sup>21</sup>; poly(A)<sup>+</sup> RNA was purified from total RNA using oligo(dT)-Latex beads (Japan Roche). Poly(A)<sup>+</sup> RNA 1.5–2.0  $\mu$ g was electrophoresed in a 1.0% agarose gel containing 1.1 M formaldehyde. RNA was transferred to Zeta Probe membranes (Bio Rad). Membranes were hybridized with radiolabelled RNA probes specific for individual mRNAs and processed for autoradiography as described<sup>22</sup>.



a control plasmid containing only the rabbit  $\beta$ -globin gene minimal promoter (pRBGP3), reporter gene activity increased significantly in cells transfected with the test plasmid pRBGP2 containing NF-E2 sites of chicken  $\beta$ -globin enhancer in front of the minimal promoter. This NF-E2 site-dependent endogenous activity was predictable because the binding sequence of NF-E2 encompasses an AP-1 site (Fig. 1a). The important result was that co-transfection of pEFmaff with the pRBGP2 reporter plasmid efficiently repressed NF-E2 site-dependent luciferase activity. In contrast, the reporter activity of pRBGP3 was only a small percentage of that of pRBGP2 and this basal activity was not affected by the effector plasmid (results not shown). As MafF lacks a canonical transactivation domain, and as the MafF homodimer can bind to the NF-E2 site of chicken  $\beta$ -globin enhancer (Fig. 2b), this result strongly suggests that MafF suppresses reporter gene activity by competing with the binding of endogenous transcription factors (probably AP-1) for the NF-E2 site in the reporter plasmid.

This suppression by MafF was completely reversed by co-transfecting pEFp45 with pEFmaff (Fig. 3), and the effect of p45 expression on the MafF repressor function was dependent on the amount of transfected pEFp45 plasmid. Similar results were obtained using a MafK-expression plasmid instead of the MafF-expression plasmid, by using quail fibroblast QT6 as recipient cells, or by using a luciferase reporter gene driven by the SV40 early promoter with the triplicated NF-E2 site (not shown). As the heterodimer of p45 and MafF has a much higher binding affinity for the NF-E2 site than the MafF homodimer, the antagonistic effect of p45 on the repressor function of MafF should be the result of transcriptional activation generated by the p45/MafF heterodimer, rather than mere displacement of the MafF homodimer from the NF-E2 sites making the NF-E2 sites accessible to intrinsic transcription factors.

The *maff* and *mafK* genes are distinctly expressed in a range of tissues in the chicken<sup>3</sup>. We investigated whether messenger RNAs for the small Maf proteins are also expressed in cells of erythroid lineage using RNA blot hybridization analysis and found that these mRNAs are expressed in erythroid (HD3; ref. 14) as well as lymphoid (MSB-1; ref. 15) cell lines (Fig. 4), supporting the idea that the Maf proteins can function in the erythroid lineage.

A protein of  $M_r$  18,000 (p18) that copurifies with p45 (ref. 2) has been proposed to create the functional form of NF-E2 by heterodimerization with p45. The p18 polypeptide is comparable in size to the small Maf proteins<sup>3</sup>, and therefore may be the product of a mouse *maf* gene homologue. The mouse MafK protein sequence is 96% identical to chicken MafK, and the mouse p45 and MafK heterodimer shows a markedly increased affinity for NF-E2 binding sequences compared with either protein alone (unpublished observation). Our results indicate that homodimers of the MafF and MafK proteins function as transcriptional repressors, and therefore suggest that expression of p45 acts as a sophisticated switch in the regulation of erythroid-specific gene expression. Negative as well as positive regulation can be elicited through the NF-E2 sites, depending on the equilibrium concentrations of p45 and of small Maf protein family members in erythroid cells. Ubiquitously expressed Maf proteins may act *in vivo* by forming heterodimers with other tissue-specific components, as with NF-E2 in erythroid cells. □

Received 21 October; accepted 23 December 1993.

- Orkin, S. H. *Cell* **63**, 665–672 (1990).
- Andrews, N. C., Erdjument-Bromage, H., Davidson, M. B., Tempst, P. & Orkin, S. H. *Nature* **362**, 722–728 (1993).
- Fujiwara, K. T., Kataoka, K. & Nishizawa, M. *Oncogene* **8**, 2371–2380 (1993).
- Nishizawa, M., Kataoka, K., Goto, N., Fujiwara, K. T. & Kawai, S. *Proc. natn. Acad. Sci. U.S.A.* **86**, 7711–7715 (1989).
- Kataoka, K., Nishizawa, M. & Kawai, S. *J. Virol.* **67**, 2133–2141 (1993).
- Kataoka, K., Noda, M. & Nishizawa, M. *Molec. cell. Biol.* **14**, 700–712 (1994).
- Mignotte, V., Elouet, J. F., Raich, N. & Romeo, P.-H. *Proc. natn. Acad. Sci. U.S.A.* **86**, 6548–6552 (1989).
- Reitman, M. & Felsenfeld, G. *Proc. natn. Acad. Sci. U.S.A.* **85**, 6267–6271 (1988).

- Moon, A. M. & Lay, T. J. *Proc. natn. Acad. Sci. U.S.A.* **87**, 7693–7697 (1990).
- Cox, T. C., Bawden, M. J., Martin, A. & May, B. K. *EMBO J.* **10**, 1891–1902 (1991).
- Kouzarides, T. & Ziff, E. *Nature* **340**, 568–571 (1989).
- O'Shea, E., Rutkowski, R., Stafford, W. F. & Kim, P. *Science* **245**, 646–648 (1989).
- Vinson, C. R., Hai, T. & Boyd, S. M. *Genes Dev.* **7**, 1047–1058 (1993).
- Beug, H., Palmieri, S., Freudestein, C., Zeutgraf, H. & Graf, T. *Cell* **28**, 907–919 (1982).
- Akiyama, Y. & Kato, S. *Biken J.* **17**, 195–116 (1974).
- Yamamoto, M. et al. *Genes Dev.* **4**, 1650–1662 (1990).
- Muller, M. M., Ruppert, S., Schaffner, W. & Matthias, P. *Nature* **336**, 544–551 (1988).
- Mizushima, S. & Nagata, S. *Nucleic Acids Res.* **18**, 5322 (1990).
- Maniatis, T., Fritsch, E. F. & Sambrook, J. *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor, New York, 1989).
- Moscovici, C. et al. *Cell* **11**, 95–103 (1980).
- Chomczynski, P. & Sacchi, N. *Analyt. Biochem.* **162**, 156–159 (1987).
- Fujita, H., Yamamoto, M., Yamagami, T., Hayashi, N. & Sassa, S. *J. biol. Chem.* **266**, 17494–17502 (1991).

ACKNOWLEDGEMENTS. We thank J. D. Engel, S. Shibahara, T. Suda and B. Roizman for critical reading of the manuscript, Y. Nabeshima for pENL, S. Nagata for pEF-BOS and T. Sasaki for his help. This work was supported in part by Grants-in-Aid from the Ministry of Education, Science and Culture, a grant from the Gono Foundation for Promotion of Medical Science, and a grant from the Uehara Memorial Foundation.

## Initial events of myelination involve Fyn tyrosine kinase signalling

Hisashi Umemori\*, Shuzo Sato†, Takeshi Yagi‡, Shinichi Aizawa‡ & Tadashi Yamamoto\*§

\* Department of Oncology, Institute of Medical Science, University of Tokyo, 4–6–1 Shirokanedai, Minato-ku, Tokyo 108, Japan

† Department of Neurology, Brain Research Institute, Niigata University, Niigata 951, Japan

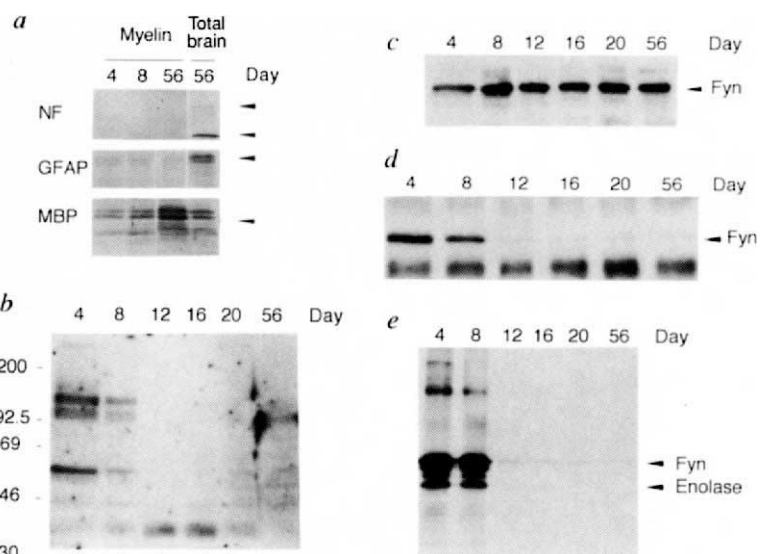
‡ Laboratory of Molecular Oncology, Tsukuba Life Science Center, Institute of Physical and Chemical Research (RIKEN), Tsukuba, Ibaraki 305, Japan

MYELIN is the lipoprotein multimembrane that functions as an insulator preventing the flow of ion currents across the axonal membrane and facilitating the conduction of nerve impulses. It is synthesized by oligodendrocytes in the central nervous system at about the time of birth in mammals<sup>1</sup>. During the initial stages of myelination, several proteins are phosphorylated on tyrosine. Among these proteins, we identified Fyn tyrosine kinase, one of the non-receptor-type tyrosine kinases of the Src family. Here we report that Fyn tyrosine kinase is activated during the initial stages of myelination and that it is associated with the large myelin-associated glycoprotein (MAG), an adhesion molecule that has been implicated in myelinogenesis<sup>2</sup>. The Fyn-large MAG association requires amino-terminal domains of Fyn that include SH2 and SH3 (Src homology domains 2 and 3). Crosslinking of large MAG with antibody induces a rapid increase in the specific activity of Fyn kinase. These results indicate that Fyn participates in the initial events of myelination as a signalling molecule downstream of large MAG; indeed, we find that *fyn*-deficient mice exhibit impaired myelination.

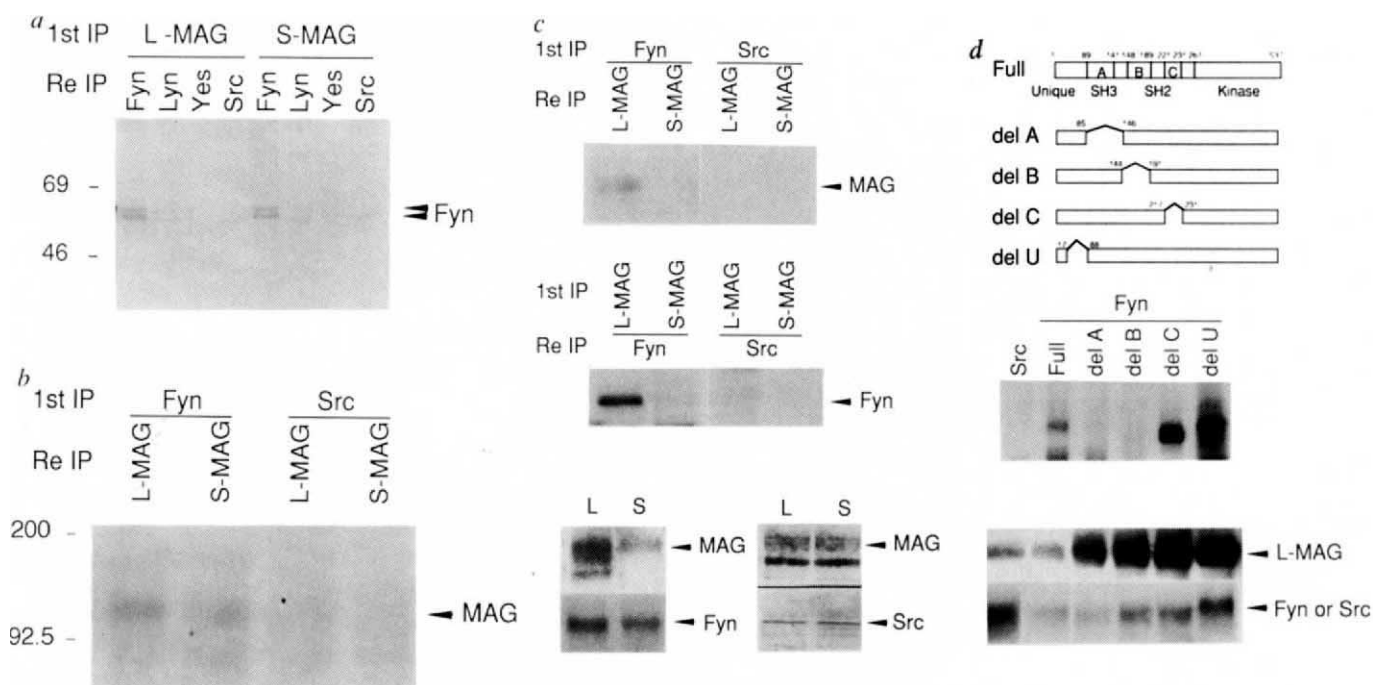
Myelin was isolated from mouse brain at 4–56 days of age (P4–P56)<sup>3</sup>. All preparations had the same degree of purity, even at the youngest age, as demonstrated by immunoblotting with antibodies specific for neurons, astrocytes and oligodendrocytes (Fig. 1a). Anti-phosphotyrosine immunoblotting of total myelin lysates revealed several distinct proteins phosphorylated at tyrosine residues during myelination (Fig. 1b). At least four of the proteins were significantly phosphorylated and three of them, including a protein of relative molecular mass 59,000 ( $M_r$  59K), showed highest phosphorylation during the initial stage of myelination (P4). As the Src-related tyrosine protein kinase (TPK) p59<sup>src</sup> is expressed in oligodendrocytes<sup>4</sup>, we investigated the pos-

§ To whom correspondence should be addressed.

**FIG. 1** Tyrosine phosphorylation and kinase activity of Fyn during myelination. **a**, Purity of the isolated myelin lysate. Immunoblotting was carried out using antibodies specific for neuron (neurofilament; NF), astrocytes (glial fibrillary acidic protein; GFAP) and oligodendrocytes (myelin basic protein; MBP). The position of each protein is indicated by an arrowhead. The  $M_r$ s of NF subunits, GFAP and MBP are 200K and 68K, 56K and 14–21.5K, respectively. **b**, Tyrosine phosphorylated proteins during myelination (P4–P56). The positions of prestained  $M_r$  markers ( $\times 10^{-3}$  Amersham) are indicated. **c**, Expression of Fyn during myelination. **d**, Tyrosine phosphorylation of Fyn during myelination. **e**, Tyrosine kinase activity of Fyn during myelination. **METHODS**. Myelin was isolated from the brain of BALB/c mice on postnatal days 4, 8, 12, 16, 20 and 56 (ref. 3). The brains were homogenized in 10 vol. (w/v) of 0.32 M sucrose, layered over 0.85 M sucrose, centrifuged at 75,000g for 30 min, and the layers of crude myelin which formed at the interface of the two sucrose solutions were collected. The crude myelin was suspended in water and centrifuged at 75,000g for 15 min. The resultant pellets were dispersed in water and centrifuged twice at 12,000g for 10 min. The pellets were suspended in 0.32 M sucrose, layered over 0.85 M sucrose, centrifuged as above, and the purified myelin was removed from the interface. The purified myelin and total adult brain were lysed in TNE buffer<sup>20</sup> and equal amounts of lysates (20  $\mu$ g protein) were subjected to SDS-PAGE followed by immunoblotting with monoclonal anti-NF, monoclonal anti-GFAP and polyclonal anti-MBP antibodies (Nichirei) (**a**), polyclonal antiphosphotyrosine (APT) antibody<sup>20</sup> (**b**), or polyclonal anti-Fyn antibody<sup>7</sup> (**c**). The lysates (200  $\mu$ g protein) at each stage were immunoprecipitated with monoclonal APT antibody (PY20; ICN) and the immunoprecipitates were resolved by 9% SDS-PAGE followed by immunoblotting for Fyn (**d**). In kinase assays, 100  $\mu$ g



protein of lysates were used in each reaction (**e**). Immune-complex kinase assays were carried out as reported<sup>20</sup>.



**FIG. 2** Association of Fyn with MAG. **a**, Co-immunoprecipitation of Fyn with large (L) and small (S) MAGs from a brain lysate. The  $M_r$  markers ( $\times 10^{-3}$ ) are indicated. When Fyn immunoprecipitates are analysed on SDS-PAGE, two Fyn bands are often detected<sup>21</sup>. The lower band (56K) may be a degradation product of the upper band (59K), but its nature is still unclear. **b**, Co-immunoprecipitation of MAGs with Fyn from a brain lysate. **c**, Association of Fyn and large MAG in cotransfected COS cells. Top panel, co-immunoprecipitation of large MAG with Fyn; middle panel, co-immunoprecipitation of Fyn with large MAG; bottom panel, immunoblotting showing that transfected large and small MAG, Fyn and Src were certainly expressed in each cotransfected COS cells. **d**, Association of mutant Fyns with large MAG in cotransfected COS cells. Top panel, schematic depiction of Fyn mutants. Residue numbers correspond to those in the amino acid sequence of Fyn. Middle panel, co-immunoprecipitation of mutant Fyns with large MAG. Bottom panel, expressions of large MAG, Src and mutant Fyns in transfected COS cells. **METHODS**. Whole brains of adult BALB/c mice were lysed in digitonin lysis buffer<sup>13</sup>. The lysates (500  $\mu$ g protein) were incubated with the first antibodies (1st IP), and the immunoprecipitates were subjected to the kinase reaction. The reaction products were dissociated from the immunoabsorbent with solubilizing buffer<sup>13</sup> and the extracts were re-immunoprecipitated with the

second antibodies (Re IP) as described<sup>13</sup>. Polyclonal antibodies directed against large and small MAG<sup>22</sup> and monoclonal antibodies against Lyn (Lyn8)<sup>20</sup>, Yes (3H9)<sup>23</sup> and Src (GD11)<sup>24</sup> were as reported. Full-length cDNAs of large and small MAG<sup>25</sup>, *fyn*<sup>26</sup> and *src*<sup>27</sup>, respectively, were cloned into the expression vector pME18S (ref. 28). COS cells were transfected with the *fyn* or *src* expression plasmid and large or small MAG expression plasmid by the calcium phosphate method<sup>29</sup>. The cells were collected 48 h after transfection and co-immunoprecipitation analysis was performed as described above. Expression of transfectants were confirmed by immunoblotting. Fyn mutants were constructed as follows: *Clal* sites were introduced into the *fyn* cDNA sequence at nucleotide positions 631, 835, 1,009, 1,147, 1,222 and 1,273, respectively, by site-directed point mutation<sup>30</sup>. The cDNAs were then digested with *Clal* and the fragments were ligated to generate a series of deletion mutants; del A was constructed by ligation of the fragment encoding amino acid residues 1–85 and the fragment encoding residues 146–537; del B consisted of residues 1–144 plus 191–537; del C consisted of residues 1–217 plus 231–537; and del U consisted of residues 1–17 plus 88–537. Cotransfection and co-immunoprecipitation were performed as described above.



sible role of Fyn kinase in myelination. The level of Fyn protein in myelin lysates was almost constant throughout the time period examined (Fig. 1c). In contrast, the level of phosphorylation on tyrosine residues in Fyn was maximal during the initial stage of myelination (P4) and declined thereafter (Fig. 1d). Our preliminary experiments suggested that the reduction in tyrosine phosphorylation is associated with dephosphorylation at Tyr 420 of Fyn (data not shown). This finding that total tyrosine phosphorylation of Fyn was downregulated might be an artefact of

poor recognition of phosphorylated Tyr 531 by the anti-phosphotyrosine antibody, if every molecule of Fyn was phosphorylated at residue 420 or 531. The kinase activity of Fyn was also highest on P4, and then decreased rapidly, as measured by *in vitro* autophosphorylation and phosphorylation of enolase (Fig. 1e). The phosphorylation on residue 420 is thought to activate Fyn kinase<sup>3</sup>. Thus we conclude that myelin Fyn kinase is phosphorylated at Tyr 420 and is most active in the initial stage of myelination.

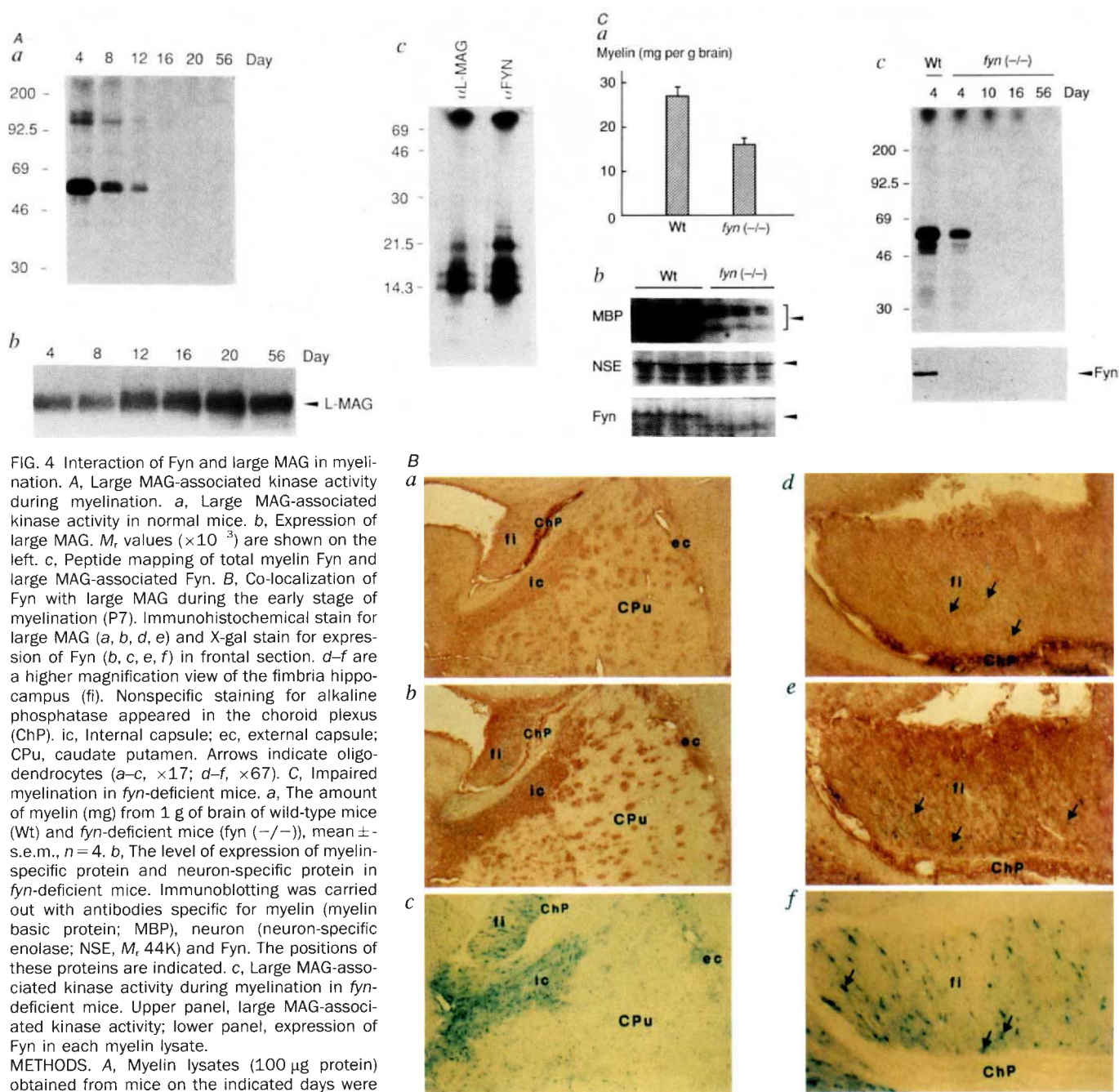


FIG. 4 Interaction of Fyn and large MAG in myelination. **A**, Large MAG-associated kinase activity during myelination. **a**, Large MAG-associated kinase activity in normal mice. **b**, Expression of large MAG.  $M_r$  values ( $\times 10^{-3}$ ) are shown on the left. **c**, Peptide mapping of total myelin Fyn and large MAG-associated Fyn. **B**, Co-localization of Fyn with large MAG during the early stage of myelination (P7). Immunohistochemical stain for large MAG (**a**, **b**, **d**, **e**) and X-gal stain for expression of Fyn (**b**, **c**, **e**, **f**) in frontal section. **d**–**f** are a higher magnification view of the fimbria hippocampus (fi). Nonspecific staining for alkaline phosphatase appeared in the choroid plexus (ChP). ic, Internal capsule; ec, external capsule; CPu, caudate putamen. Arrows indicate oligodendrocytes (**a**–**c**,  $\times 17$ ; **d**–**f**,  $\times 67$ ). **C**, Impaired myelination in *fyn*-deficient mice. **a**, The amount of myelin (mg) from 1 g of brain of wild-type mice (Wt) and *fyn*-deficient mice (*fyn* (-/-)), mean  $\pm$  s.e.m.,  $n = 4$ . **b**, The level of expression of myelin-specific protein and neuron-specific protein in *fyn*-deficient mice. Immunoblotting was carried out with antibodies specific for myelin (myelin basic protein; MBP), neuron (neuron-specific enolase; NSE,  $M_r$  44K) and Fyn. The positions of these proteins are indicated. **c**, Large MAG-associated kinase activity during myelination in *fyn*-deficient mice. Upper panel, large MAG-associated kinase activity; lower panel, expression of Fyn in each myelin lysate.

**METHODS.** **A**, Myelin lysates (100  $\mu$ g protein) obtained from mice on the indicated days were incubated with anti-large MAG antibody and the immunoprecipitates were subjected to immune-complex kinase assay or immunoblotting. Representative  $^{32}$ P-labelled bands corresponding to Fyn in Fig. 1d and the 59K phosphoprotein in A,a (P4) were isolated and subjected to partial proteolytic peptide analysis using *Staphylococcus aureus* V8 protease. The mapping pattern is identical to that of Fyn reported previously<sup>12</sup>. **B**, Frontal  $+/-fyn^2$  heterozygous frozen sections were processed for large MAG immunohistochemistry (Fast red)<sup>22</sup> and/or incubated with X-gal reaction mixture so that the Fyn-fused LacZ stained blue<sup>18</sup>. **C**, Myelin was isolated from the

brain of 4-week-old wild-type and *fyn*-deficient mice as described for Fig. 1 ( $n = 4$ ). Isolated myelin was freeze-dried and weighed. The mass of myelin per gram brain (w/w) is shown (**a**). Equal amounts of brain lysates (100  $\mu$ g protein) obtained from wild-type and *fyn*-deficient mice were subjected to immunoblotting with polyclonal anti-MBP, polyclonal anti-NSE (Nichirei) and polyclonal anti-Fyn antibodies (**b**). Immune-complex kinase assay was as described for **A** (**c**).

Axon–glia interaction is essential for the formation of a mature, fully differentiated myelinated nerve fibre. Previous observations<sup>6,9</sup>, including the fact that monoclonal anti-MAG antibodies block neuron–oligodendrocyte and oligodendrocyte–oligodendrocyte adhesions<sup>9</sup>, suggest that among the recently identified cell-surface constituents of myelinating glial cells, MAG proteins are the best candidates for mediating this interaction. Two MAGs, large and small, are present in the central nervous system (CNS), and large MAG, which has a tyrosine phosphorylation site in its cytoplasmic domain<sup>10</sup>, is predominant during the early and active stages of CNS myelination<sup>1</sup>. As several Src-related TPKs associate functionally with cell-surface receptors<sup>11–15</sup>, we examined whether Fyn is associated with MAGs. From total brain lysates, Fyn was co-immunoprecipitated with both forms of MAG (Fig. 2a), and conversely, MAGs were co-immunoprecipitated with Fyn (Fig. 2b). Among the Src-related TPKs expressed in the brain<sup>4,16</sup>, Fyn showed the strongest association, Lyn, Yes and Src showing much weaker association (Fig. 2a). Densitometric analysis revealed that the recovery of Fyn in anti-MAG immunoprecipitates from myelin lysate was about 5% and the recovery of MAG in anti-Fyn immunoprecipitates was 1–2% (data not shown).

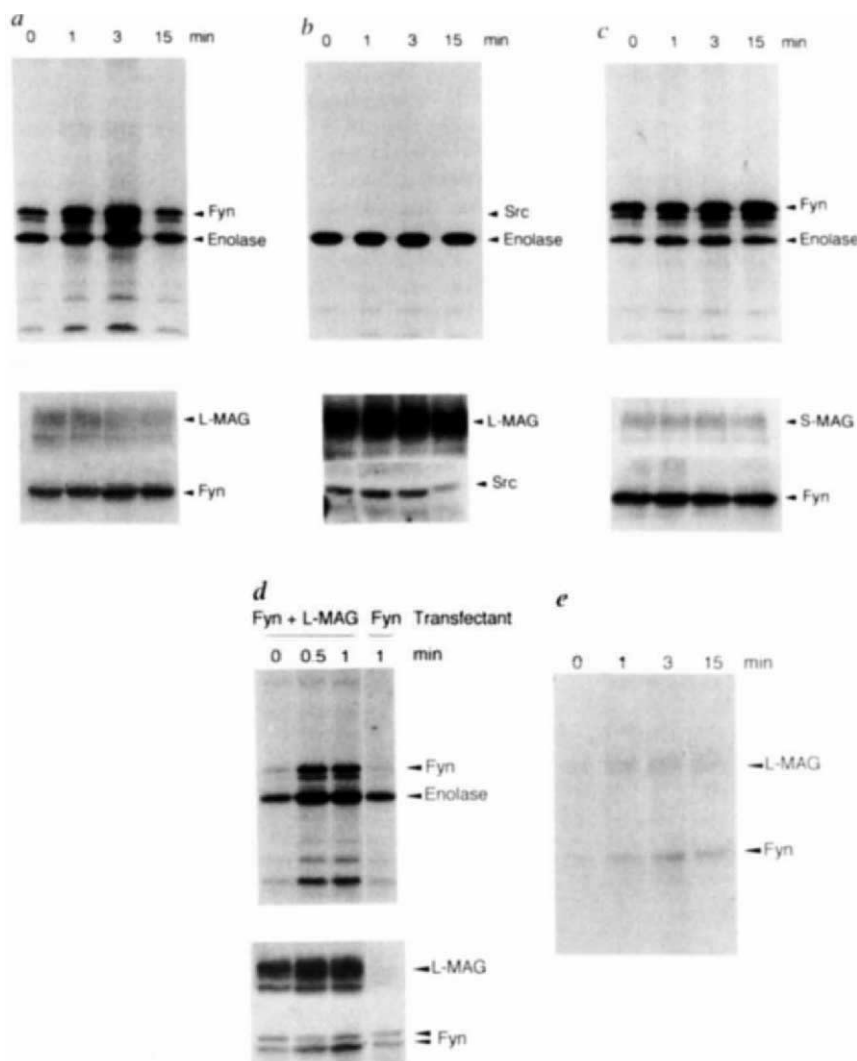
To gain further insight into the nature of the interaction between Fyn and MAGs, we used a transient complementary DNA expression system in COS cells. COS cells were cotransfected with large or small MAG and *fyn* or *src* expression plasmids. Co-immunoprecipitation analysis showed that Fyn was associated with large MAG, whereas no association was detected

between Fyn and small MAG or Src and MAGs (Fig. 2c). Therefore, the association of Fyn and large MAG does not require any other brain-specific components. In contrast, the association of Fyn and small MAG could be mediated by other molecules in the brain. To localize the large MAG interaction sites on Fyn, we generated a series of Fyn mutants lacking, respectively, the unique region and Src homology sequences, that is, SH3 (region A) and SH2 (region B or C). All the Fyn mutants retained tyrosine kinase activity (data not shown). Cotransfection of the mutant *fyn* cDNAs with the large MAG cDNA followed by co-immunoprecipitation analysis revealed that mutants lacking SH3 or the B region of SH2 did not interact with large MAG (Fig. 2d). Thus, both SH3 and the B region in SH2 of Fyn are involved in its association with large MAG. Like the association between activated Src and phosphoprotein p110 (ref. 17), both SH2 and SH3 of Fyn would contribute to its stable association with MAG; this may cause the cytoskeleton to form compact myelin.

To test whether binding of ligand to the extracellular domain of large MAG results in intracellular signalling involving Fyn, we crosslinked the extracellular domain of large MAG with specific antibody. In cotransfected COS cells, crosslinking of large MAG with anti-MAG monoclonal antibody (the epitope recognized by this antibody is probably localized in the protein part of the molecule<sup>9</sup>) and goat anti-mouse IgG resulted in a rapid (within 1 min) and significant (three- to fivefold) increase in the TPK activity of Fyn (both autophosphorylation and phosphorylation of enolase), with no change in the amounts of Fyn and large MAG (Fig. 3a); large MAG crosslinking had no effect

FIG. 3 Activation of Fyn after large MAG crosslinking. **a**, Time course of change in kinase activity of Fyn after large MAG crosslinking. Upper panel, kinase activity of Fyn; lower panel, expressions of large MAG and Fyn. **b**, Time course of change in kinase activity of Src after large MAG crosslinking. Upper panel, kinase activity of Src; lower panel, expressions of large MAG and Src. **c**, Time course of change in kinase activity of Fyn after small MAG crosslinking. Upper panel, kinase activity of Fyn; lower panel, expressions of small MAG and Fyn. **d**, Time course of change in kinase activity of Fyn after large MAG crosslinking by polyclonal anti-MAG antibody. Upper panel, kinase activity of Fyn; lower panel, expressions of large MAG and Fyn. When Fyn was transfected alone, activation of Fyn kinase was not induced by crosslinking. **e**, Time courses of changes in tyrosine phosphorylation (APT immunoblotting) of large MAG and Fyn after large MAG crosslinking.

**METHODS.** COS cells ( $1.5 \times 10^6$ ) were transfected with *fyn* and large MAG (**a**, **d**, **e**), *src* and large MAG (**b**), or *fyn* and small MAG (**c**) expression plasmids by the calcium phosphate method and incubated for 48 h. The cells were then incubated on ice with monoclonal anti-MAG antibody<sup>9</sup> ( $2.5 \mu\text{g ml}^{-1}$ ; Boehringer) for 30 min, washed and incubated at  $37^\circ\text{C}$  with goat anti-mouse IgG ( $10 \mu\text{g ml}^{-1}$ ; Organon Teknika) for the indicated times (**a**–**c**, **e**). Polyclonal anti-MAG antibody was added directly to the medium and incubated at  $37^\circ\text{C}$  for the indicated times (**d**). The cells were solubilized in TNE buffer, and equal amounts of lysates (500  $\mu\text{g}$  protein) were subjected to *in vitro* kinase assay or immunoblotting.





on Src tyrosine kinase activity (Fig. 3b). Crosslinking of small MAG had no appreciable effect on Fyn kinase activity (Fig. 3c). We therefore conclude that the crosslinking of large MAG specifically activates Fyn tyrosine kinase. Activation of Fyn kinase was also observed on crosslinking of large MAG using polyclonal anti-MAG antibody obtained from rabbits immunized with purified MAG (Fig. 3d). The time courses of tyrosine phosphorylation of Fyn and large MAG after the crosslinking paralleled the change in Fyn kinase activity (Fig. 3e).

We showed that the crosslinking of large MAG activates Fyn kinase in cotransfected COS cells. To determine how its activation is reflected in myelinogenesis, we examined large MAG-associated kinase activity during myelination (Fig. 4A). Immunoprecipitation of large MAG revealed that the activity of large MAG-associated kinase was highest during the initial stage of myelination (P4) and decreased thereafter (Fig. 4A, a and b), like the kinase activity of total myelin Fyn (Fig. 1e). Evaluation of the co-immunoprecipitating kinase by peptide mapping revealed that the large MAG-associated kinase was Fyn (Fig. 4A, c).

To confirm the interaction of Fyn and large MAG *in vivo*, we examined the localization of Fyn and large MAG during brain development (Fig. 4B). By using  $+/-fyn^c$  heterozygous mice, in which the unique region of the *fyn* gene on one of the alleles was fused to the bacterial *lacZ* gene in-frame, expression of Fyn was followed by LacZ expression<sup>18</sup>. Large MAG expression was monitored by immunohistochemistry. During the early stage of myelination (P7), Fyn and large MAG showed striking co-localization in fibre tracts and oligodendrocytes (Fig. 4B), suggesting their interactions during early myelinogenesis. We then analysed whether homozygous *fyn*-deficient mice could form myelin. Electron microscopic analysis revealed that the *fyn*-deficient mice formed myelin (data not shown), which is consistent with a previous report<sup>19</sup>. However, the amount of myelin per brain (w/w) of *fyn*-deficient mice was about 50–60% of that of age-matched wild-type mice (Fig. 4C, a). In addition, immunoblotting of brain lysates with antibodies specific for neuron (neuron-specific enolase) and myelin (myelin basic protein) revealed that 50% less myelin was present in *fyn*-deficient mice than in wild-type mice, whereas there was virtually no difference in the amount of neuron between wild-type and *fyn*-deficient mice (Fig. 4C, b). These findings indicate that myelination is partially impaired in *fyn*-deficient mice. Moreover, we detected large MAG-associated kinase in *fyn*-deficient mice, which might compensate for Fyn function (Fig. 4C, c). The kinase may belong to the Src family, and was active during the early stages of myelination, as is Fyn kinase in normal mice (Fig. 4A, a, C, c). The kinase activity was much weaker than that of normal mice, which would explain the inefficient myelination.

We propose that the interaction of large MAG with its ligand on neuronal axons generates a biological signal that involves the rapid activation of Fyn at the beginning of myelination. This signalling may be an important part of myelinogenesis. □

16. Sudol, M., Grant, S. G. N. & Maisonpierre, P. C. *Neurochem. Int.* **22**, 369–384 (1993).
17. Koch, C. A., Anderson, D., Moran, M. F., Ellis, C. & Pawson, T. *Science* **252**, 668–674 (1991).
18. Yagi, T. et al. *Nature* **366**, 742–745 (1993).
19. Grant, S. G. N. et al. *Science* **258**, 1903–1910 (1992).
20. Yamanashi, Y. et al. *Proc. natn. Acad. Sci. U.S.A.* **89**, 1118–1122 (1992).
21. Okada, M., Nada, S., Yamanashi, Y., Yamamoto, T. & Nakagawa, H. *J. biol. Chem.* **266**, 24249–24252 (1991).
22. Fujita, N. et al. *J. Neurochem.* **55**, 1056–1059 (1990).
23. Sukeyama, J. et al. *Oncogene* **5**, 611–614 (1990).
24. Parsons, S. J., McCarley, D. J., Raymond, V. W. & Parsons, J. T. *J. Virol.* **59**, 755–758 (1986).
25. Fujita, N. et al. *Biochem. biophys. Res. Commun.* **165**, 1162–1169 (1989).
26. Semba, K. et al. *Proc. natn. Acad. Sci. U.S.A.* **83**, 5459–5463 (1986).
27. Tanaka, A. et al. *Molec. cell. Biol.* **7**, 1978–1983 (1987).
28. Takebe, Y. et al. *Molec. cell. Biol.* **8**, 466–472 (1988).
29. Turner, J. M. et al. *Cell* **60**, 755–765 (1990).
30. Kunkel, T. A. *Proc. natn. Acad. Sci. U.S.A.* **82**, 488–492 (1985).

ACKNOWLEDGEMENTS. We thank M. Sudol for critical reading of the manuscript; Y. Kadowaki and Y. Yamanashi for discussion; J. Fujimoto, E. Suzuki and K. Hirose for microscopic analyses; Y. Shigetani for histochemical analyses; Y. Fukui for GD11; A. Tanaka for *src* cDNA; M. Takeuchi for construction of the Fyn mutants; and N. Fusaki and S. Nagasawa for care of mice.

## Glioblastoma growth inhibited *in vivo* by a dominant-negative Flk-1 mutant

Birgit Millauer\*, Laura K. Shawver†, Karl H. Plate‡, Werner Risau‡ & Axel Ullrich\*§

\* Department of Molecular Biology, Max-Planck-Institut für Biochemie, Am Klopferspitz 18A, 82152 Martinsried, Germany

† SUGEN Inc., 515 Galveston Drive, Redwood City, California 94063, USA

‡ Department of Molecular Cell Biology, Max-Planck-Institut für physiologische und klinische Forschung, W. G. Kerckhoff Institut, 61231 Bad Nauheim, Germany

ANGIOGENESIS, the sprouting of capillaries from pre-existing blood vessels, is a fundamental process in the formation of the vascular system during embryonic development. In adulthood, angiogenesis takes place during corpus luteum formation and in pathological conditions such as wound healing, diabetic retinopathy, and tumorigenesis. Vascularization is essential for solid tumour growth and is thought to be regulated by tumour cell-produced factors, which have a chemotactic and mitogenic effect on endothelial cells<sup>1–4</sup>. Vascular endothelial growth factor (VEGF), a homodimeric glycoprotein of relative molecular mass 45,000, is the only mitogen, however, that specifically acts on endothelial cells, and it may be a major regulator of tumour angiogenesis *in vivo*<sup>5,6</sup>. Its expression has been shown to be upregulated by hypoxia, and its cell-surface receptor, Flk-1, is exclusively expressed in endothelial cells<sup>7,8</sup>. Here we investigate the biological relevance of the VEGF/Flk-1 receptor/ligand system for angiogenesis using a retrovirus encoding a dominant-negative mutant of the Flk-1/VEGF receptor to infect endothelial target cells *in vivo*, and find that tumour growth is prevented in nude mice. Our results emphasize the central role of the Flk-1/VEGF system in angiogenesis in general and in the development of solid tumours in particular.

We recently reported the endothelial cell-specific expression of Flk-1 during the development of the mouse vascular system and its function as a receptor for VEGF<sup>7</sup>. To investigate the relevance of Flk-1 as a regulator of endothelial cell growth and angiogenesis, we used a solid-tumour growth model and a dominant-negative strategy that allows interference with receptor tyrosine kinase-mediated signal transduction by formation of inactive dimers after introduction of a signalling-defective receptor mutant into the target cell<sup>9–12</sup>. For delivery of this genetic information, we used replication-defective recombinant retroviruses, which were based on elements of the mouse sarcoma virus (MSV) genome and encoded the complete wild-type Flk-

Received 7 June; accepted 31 December 1993.

1. Duncan, I. D., Skoff, R. P. & Colman, D. (eds) *Myelination and Dysmyelination* (New York: Academy of Sciences, New York, 1990).
2. Lemke, G. *Neuron* **1**, 535–543 (1988).
3. Norton, W. T. & Poduslo, S. E. *J. Neurochem.* **21**, 749–757 (1973).
4. Umemori, H. et al. *Molec. Brain Res.* **16**, 303–310 (1992).
5. Cooper, J. A. in *Peptides and Protein Phosphorylation* (ed. Kemp, B.) 85–113 (CRC Press, Boca Raton, 1990).
6. Salzer, J. L., Holmes, W. P. & Colman, D. R. *J. Cell Biol.* **104**, 957–965 (1987).
7. Sternberger, N. H., Quarles, R. H., Itoyama, Y. & Webster, H. de F. *Proc. natn. Acad. Sci. U.S.A.* **76**, 1510–1514 (1979).
8. Johnson, P. W. et al. *Neuron* **3**, 377–385 (1989).
9. Poltrak, M. et al. *J. Cell Biol.* **105**, 1893–1899 (1987).
10. Edwards, A. M. et al. *Molec. cell. Biol.* **8**, 2655–2658 (1988).
11. Veillette, A., Bookman, M. A., Horak, E. M. & Bolen, J. B. *Cell* **55**, 301–308 (1988).
12. Samelson, L. E., Phillips, A. F., Luong, E. T. & Klausner, R. D. *Proc. natn. Acad. Sci. U.S.A.* **87**, 4358–4362 (1990).
13. Yamanashi, Y., Kakiuchi, T., Mizuguchi, J., Yamamoto, T. & Toyoshima, K. *Science* **251**, 192–194 (1991).
14. Eisman, E. & Bolen, J. B. *Nature* **355**, 78–80 (1992).
15. Hatakeyama, M. et al. *Science* **252**, 1523–1528 (1991).

§ To whom all correspondence should be addressed.

1/VEGF-R (LX Flk-1) or the dominant-negative mutant LX Flk-1 TM under the transcriptional control of the MSV 3' long terminal repeat (LTR)<sup>13</sup>. The Flk-1 TM mutant lacked 561 C-terminal amino acids of the intracellular kinase domain, but retained transmembrane domain sequences and 23 residues of the cytoplasmic portion of the receptor. As a control, we used a virus mediating the expression of an analogously mutated form of the colony-stimulating factor-1 receptor/*c-fms*.

The capability of Flk-1 TM to form signalling-incompetent heterodimers with the 180K wild-type Flk-1 was demonstrated by coprecipitation of the truncated 130K receptor mutant with an antibody against the C terminus of the intact receptor from lysates of COS cells transiently expressing both forms. As the antibody could not recognize Flk-1 TM, coprecipitation was a result of heterodimerization (Fig. 1a). The dominant-negative potential of Flk-1 TM was first examined by measuring its influence on the mitogenic response of Flk-1-expressing NIH3T3 cells after superinfection with the Flk-1 TM virus. As shown in Fig. 1b, [<sup>3</sup>H]thymidine incorporation in the 3T3 Flk-1 cell line was maximally stimulated by 500 pM VEGF, with a half-maximal effective concentration (EC<sub>50</sub>) of ~100 pM. After superinfection with the Flk-1 TM virus, the Flk-1/VEGF-mediated mitogenic response was dramatically suppressed, even at a ligand concentration of 5 nM. Although 3T3 Flk-1/Flk-1 TM cells expressed wild-type Flk-1 at levels equal to the parental line, they displayed, owing to overexpression of Flk-1 TM, a 6-fold increase of cell-surface receptors, as determined by <sup>125</sup>I-labelled VEGF binding (not shown). These results were extended by Flk-1 TM virus-induced suppression of Flk-1 transforming activity (not shown) and showed not only that mutant and wild-type Flk-1 physically associated, but also that this interaction generated signalling-incompetent heterodimers. The dominant-negative inhibitory effect that was achieved at a fivefold excess of Flk-1 TM could not be overcome by a 50-fold ligand excess relative to the EC<sub>50</sub> value for mitogenic activation. Moreover, Flk-1 TM did not interfere with the signal transduction of the  $\alpha$ - and  $\beta$ -PDGF receptors, demonstrating the specificity of its dominant-negative action (data not shown).

To investigate whether Flk-1 function is essential for promotion of tumour angiogenesis, we used C6 rat glioblastoma cells, which produce aggressive subcutaneous tumours in nude mice. *In vitro*, C6 cells secrete VEGF, which is further induced by hypoxia<sup>6,8</sup>. *In vivo*, intracerebral and subcutaneous transplants of C6 cells into syngeneic rats led to upregulation of VEGF messenger RNA in palisading glioblastoma cells surrounding necroses and an increase of Flk-1/VEGF receptor levels in tumour endothelial cells<sup>8</sup>, characteristics that make this model indistinguishable from human glioblastoma. Available data suggest that tumour-produced VEGF exerts its chemotactic, mitogenic and possibly differentiation-inducing activities on endothelial cells through the interaction with Flk-1, which results eventually in the establishment of a vascular system in the growing tumour. In case of such a prevalent and crucial role, dominant-negative inhibition of Flk-1 function in endothelial cells surrounding the emerging tumour should prevent its vascularization and consequently suppress tumour growth and disease progression.

To test this model, C6 glioblastoma cells were implanted subcutaneously in nude mice, either alone or with different relative amounts of GP+E86 cells producing comparable titres ( $1 \times 10^6$  CFU ml<sup>-1</sup>) of recombinant retrovirus. Any direct influence of the retroviruses on the growth of the tumour cells was excluded by growing C6 cells in conditioned media of the different retrovirus-producing cell lines, without any effect on their growth behaviour (data not shown). At regular intervals after implantation, the growth of developing tumours was measured. After about three weeks, mice were killed and their tumours resected for pathological and histological examination. A comparison of the growth curves from the different experimental groups is shown in Fig. 2a. The group co-implanted with the

control virus-producer cell line, NT *c-fms* TM, showed the same growth behaviour as the implant of C6 cells alone. This was also true for a group co-implanted with GP+E86 cells expressing wild-type Flk-1. In contrast, the group inoculated with retrovirus-producing cells inducing the expression of the Flk-1 dominant-negative deletion mutant, LX Flk-1 TM, had tumour growth dramatically inhibited. The effect was dose-dependent, with a maximum achieved when the virus-producing cells were in 20-fold excess over the tumour cells. There was similar inhibition of C6 glioma tumour growth when Flk-1 TM virus particle-

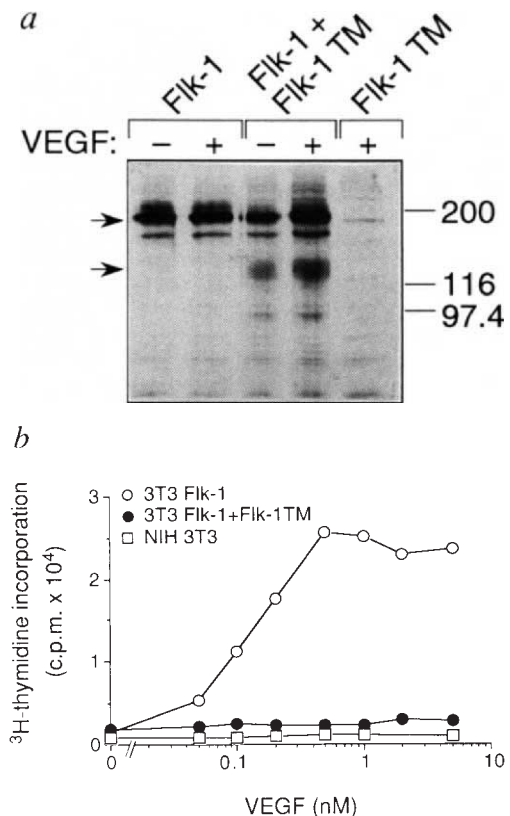


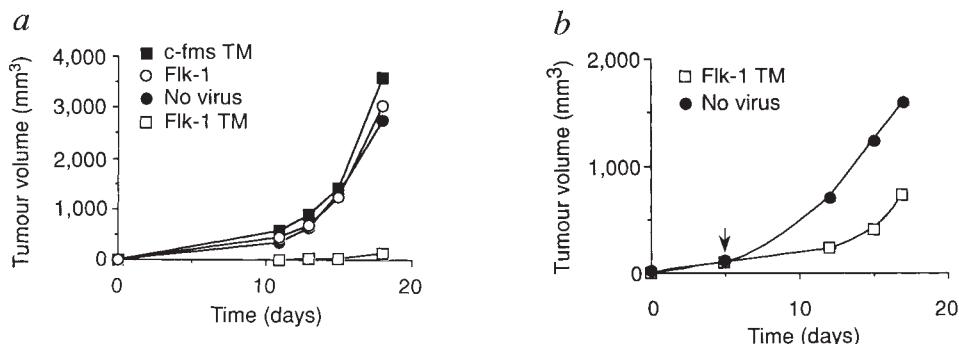
FIG. 1 Dominant-negative inhibition of Flk-1 activity by formation of signalling-defective heterodimers with Flk-1 TM. **a**, COS7 cells were transfected with cytomegalovirus (CMV) promoter-based Flk-1 and Flk-1 TM cDNA expression vectors<sup>17,18</sup>. [<sup>35</sup>S]methionine-labelled cells were treated with 1 nM VEGF for 3 h at 4 °C, lysed, and analysed by immunoprecipitation with antibody CT128 against the C terminus of Flk-1 (ref. 7) and polyacrylamide gel electrophoresis. M, markers are indicated on the right. **b**, VEGF-stimulated [<sup>3</sup>H]thymidine incorporation of NIH 3T3 cells expressing wild-type Flk-1 alone (3T3 Flk-1) or wild-type Flk-1 and a 5-fold excess of Flk-1 TM (3T3 Flk-1 + Flk-1 TM). Cells were grown to confluence in 12-well dishes (Costar), starved for 2 days in DMEM/0.5% FCS, stimulated with different concentrations of VEGF for 18 h, then incubated with 1  $\mu$ Ci [<sup>3</sup>H]thymidine per well for 4 h. Incorporation into DNA was determined as described<sup>10</sup>.

**METHODS.** Generation of recombinant retroviruses: The retroviral expression vector pLXSN has been described<sup>13</sup>. LX Flk-1 represents the entire coding region including 5' and 3' flanking sequences (4,447 bp) of Flk-1 in pLXSN. LX Flk-1 TM was obtained by ligating the 5' 2,619 bp of Flk-1 cDNA encoding amino acids 1–806 as a *Clal*/*Bam*HI fragment to a *Bgl*II/*Hpa*I linker, thereby designing a stop codon 23 amino acids behind the transmembrane region (5' GTC ATG GAT Ctt cgt taa 3'). In a second step, the *Clal*/*Hpa*I-fragment was subcloned into the *Clal*/*Hpa*I site of the pLXSN vector. Stable GP+E86 cell lines producing ecotropic retroviruses expressing the wild-type and mutated receptor constructs were generated as described<sup>10</sup>. Titres were determined by infecting NIH3T3 cells with serial dilutions of virus supernatants followed by G418 selection. Titres of the clones used were  $\sim 1 \times 10^6$  ml<sup>-1</sup>. Cell culture: GP+E86 cells were cultured in 225 cm<sup>2</sup> flasks (Nunc) in DMEM/4500 mg l<sup>-1</sup> glucose (Gibco) supplemented with 10% fetal calf serum.



FIG. 2 a, Suppression of C6 glioblastoma tumour growth by dominant-negative inhibition of Flk-1. C6 cells ( $5 \times 10^5$ ) were subcutaneously implanted in nude mice either alone or with  $1 \times 10^7$  retrovirus-producing cells. Viruses used were LX Flk-1 and LX Flk-1 TM, viruses inducing the expression of wild-type Flk-1 or a dominant-negative Flk-1 mutant under the control of the MLV 3' LTR, and NT c-fms TM, a control virus encoding a dominant-negative c-fms mutant analogous in structure to Flk-1 TM under the control of the thymidine kinase promoter.

When the first tumours appeared, their volumes were measured every 2–3 days. At the end of the experiment, mice were killed by cervical dislocation and the tumours were resected for pathological and histological examination. The dominant-negative effect was maximal (96%) at 1:20 ratio of C6 tumour cells to virus-producing cells. The inhibition was calculated by comparison of the tumour volumes of the groups co-implanted with LX Flk-1 TM virus-producing cells and control groups. Four animals per group were used, which was confirmed in an independent set of experiments with 8 animals per group. b, Inhibition of C6 glioblastoma tumour growth by localized injection of retroviral supernatants. C6 cells ( $1 \times 10^6$ ) were subcutaneously implanted in nude



containing producer cell supernatants were injected five days after implantation of  $10^6$  tumour cells, indicating that even established tumours may be suppressed by Flk-1 dominant-negative action (Fig. 2b). The transient nature of the effect, however, suggests that for lasting inhibition a continuous supply of high titres of virus is necessary.

To confirm that the inhibition of the C6 glioblastoma growth was caused by dominant-negative action of the retrovirally expressed constructs on endothelial cells, the tumours were resected and analysed. Comparison of whole-mount specimens revealed striking differences: whereas the surface of the control tumours was reddish, as expected for well-vascularized tissue, the inhibited cell implants were very pale (Fig. 3a). Histological staining of frozen sections revealed that the control tumours consisted of a homogeneous mass of viable cells (Fig. 3b, d). Only very few and small necroses could be detected. In contrast, the much smaller, growth-inhibited tumour cell implants had an onion-like histological appearance, which was characterized by different tissue layers: a large, central necrosis was surrounded by a dense layer of viable tumour cells (Fig. 3c, e). Invasion of this tumour had not progressed, as evidenced by the presence of natural structures of the skin, such as the muscular cell layer (Fig. 3f).

The distribution of capillaries and blood vessels in the tissue specimens was determined by incubating frozen tissue sections with a rat monoclonal antibody specific for the endothelial cell adhesion antigen PECAM<sup>14</sup>. The tumours co-implanted with the control virus-producing cells displayed the pattern of capillaries and vessels expected for well-vascularized tissue (Fig. 4a), whereas the growth-inhibited tumour cell implant showed a large central tumour cell necrosis, which was surrounded by a layer of viable tumour cells lacking blood vessels or capillaries (Fig. 4b). The tumour cells in this layer showed a higher cell density than the control tumour, suggestive of a reduction in tumour-induced oedema. As VEGF seems to induce vascular permeability *in vivo*, and is also designated vascular permeability factor, inhibition of VEGF/Flk-1 interaction may inhibit tumour-associated oedema.

Flk-1 expression in the proliferating endothelial cells of the tumour was confirmed by *in situ* hybridization of tissue sections with a <sup>35</sup>S-labelled Flk-1-specific antisense complementary DNA probe<sup>7</sup>: distribution was the same as immunostaining with endothelial cell-specific antibodies, indicating that proliferating endothelial cells expressed Flk-1 (Fig. 4c). *In situ* hybridization with a neomycin-resistance gene (*neo*<sup>r</sup>) antisense probe confirmed the presence of retroviral sequences. As shown in Fig. 4d, the entire

mice. At day 5 after implantation (arrow), growing tumours were injected with 100 µl retroviral supernatants (corresponding to  $\sim 10^5$  virus particles) into the site of tumour implantation. Tumour volumes were measured twice a week.

**METHODS.** Recombinant retroviruses expressing LX Flk-1 and LX Flk-1 TM were determined as for Fig. 1. For generation of NT c-fms TM, a stop codon was introduced behind amino acid 541 downstream of the transmembrane region of the c-fms cDNA using the oligonucleotide 5'-TTG TAC AAG TAT AAG TAG TAG CCC AGG TAC CAG-3'. The mutated receptor was subcloned in the retroviral expression vector pNTK2 (ref. 19).

Flk-1 dominant-negative-inhibited tumour, consisting of the retrovirus-producing and infected cells, was *neo*<sup>r</sup>-positive, which matched the result obtained with a Flk-1-specific probe (not shown). The morphology of tumours that had been co-implanted with control virus-producing cells was very similar, but the virus-producing cells were extensively infiltrated by infected tumour cells. In these tumours, which contained many capillaries and blood vessels, *neo*<sup>r</sup>-positive signals were also found in endothelial cells (not shown).

Experiments with other tumour types (not shown) support these results from the C6 glioblastoma model and indicate that solid tumour growth can be inhibited by prevention of angiogenesis, a process normally regulated by VEGF, which when secreted by tumour cells attracts and stimulates Flk-1-positive vascular endothelial cells in a paracrine fashion.

The ability to block VEGF-mediated signal transduction with the Flk-1 dominant-negative mutant emphasizes the central role of the Flk-1/VEGF system in tumour angiogenesis and further supports the contention that this system is the major regulator of angiogenesis and vasculogenesis in the whole organism. Our results are consistent with the observation that neutralizing antibodies against VEGF suppress tumour growth *in vivo*<sup>15</sup>, and suggest that the function of the alternative VEGF receptor, Flt-1 (ref. 16), is either not relevant in this system or may be blocked by heterodimer formation with the Flk-1 TM mutant. Moreover, with the identification of Flk-1 as a regulator molecule that is specific and crucially important for solid tumour progression, a variety of strategies for the development of clinical applications can be considered. For glioblastoma, a tumour with generally poor prognosis and resistance to all available therapies, retrovirus-mediated gene therapy may be advantageous, because non-mitotic brain tissues such as neurons, glia and quiescent endothelial cells would not be infected. In fact, preliminary experiments revealed that even treatment of established C6 glioblastoma tumours with Flk-1 TM retrovirus particles significantly inhibited tumour progression (Fig. 2b). Although further investigation will be necessary to develop this laboratory model into a clinically applicable gene therapy, the Flk-1 receptor system is a promising target for the development of specific anti-tumour drugs. □

Received 20 September; accepted 21 December 1993.

1. Folkman, J. *J. natn Cancer Inst.* **82**, 4–6 (1990).
2. Folkman, J., Watson, K., Ingber, D. & Hanahan, D. *Nature* **339**, 58–61 (1989).
3. Risau, W. *Progr. Growth Factor Res.* **2**, 71–79 (1990).
4. Folkman, J. & Klagsbrun, M. *Science* **235**, 442–447 (1987).

FIG. 3 Features of Flk-1 dominant-negative-inhibited tumours. *a*, Resected specimen: control tumours are on the left, an inhibited tumour is on the right. Black dots indicate the borderline between implanted cell-derived (T) and surrounding normal skin tissue (S). *b*, Frozen section of control tumour stained with toluidine blue (low magnification), showing viable tumour cell mass without visible necroses. *d*, Haematoxylin/eosin-stained section of the same tumour at higher magnification. Red staining (arrows) indicates the presence of blood vessels. *c*, Toluidine blue, and *e*, *f*, haematoxylin/eosin-stained sections of the Flk-1 dominant-negative inhibited cell implant. A large central necrosis (N) surrounded by a layer of viable tumour cells and the muscular layer (M) of the skin is visible. Morphometric examination of the tumours revealed necrotic areas of 3% in the control tumour and 27% in the inhibited tumour cell implant. These features were evident in several tumours of all groups.

**METHODS.** Tumour specimens were embedded in tissue-tek (Miles) and frozen in liquid nitrogen. 10- $\mu$ m-thick sections were cut with a cryostat (Leica) and stained histologically with toluidine blue or haematoxylin/eosin. Bar represents 1 mm (*b*, *c*) or 400  $\mu$ m (*d*, *e*, *f*).

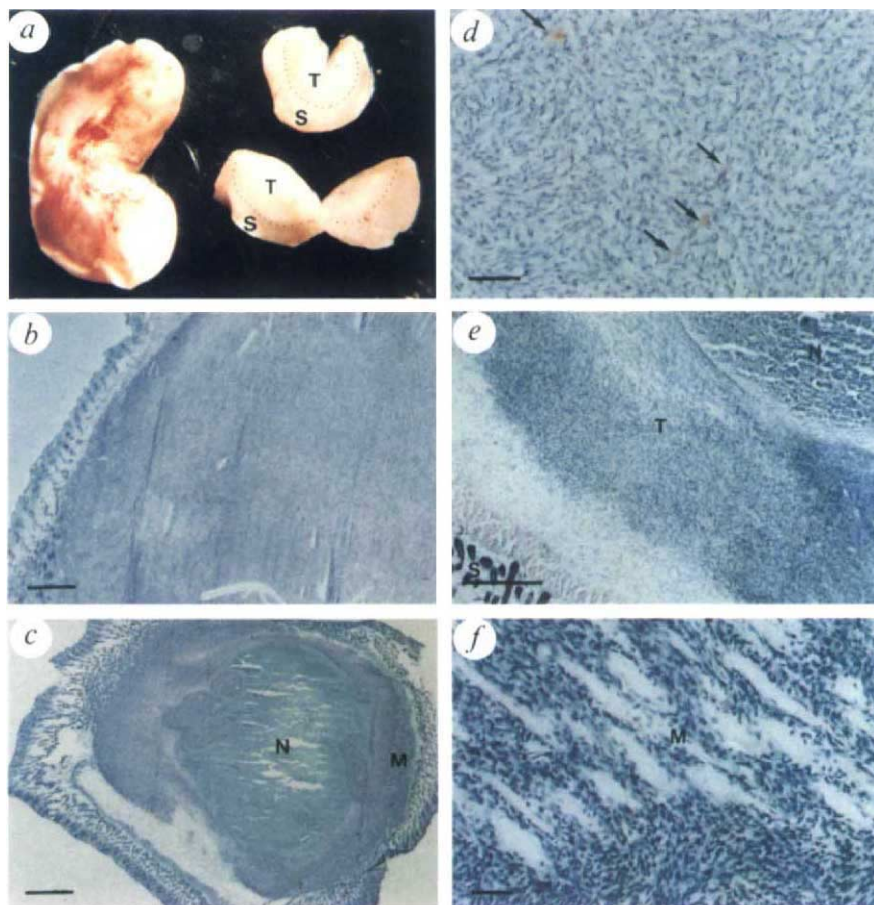
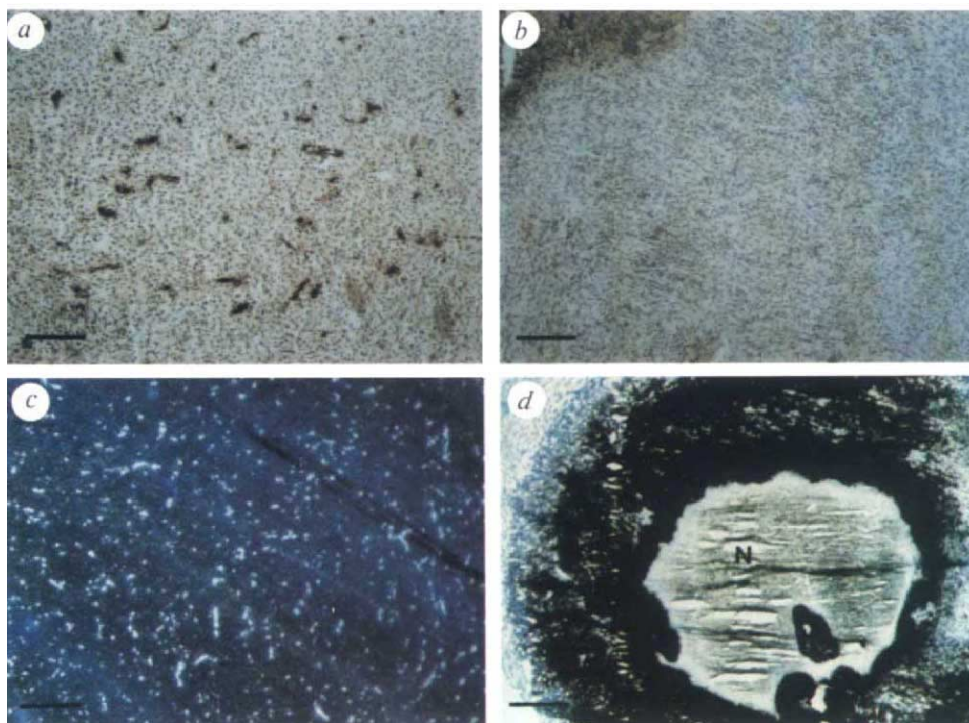


FIG. 4 *In situ* hybridization and immunohistochemical staining of tissue sections. Sections of control tumour (*a*) and Flk-1 dominant-negative inhibited tumour (*b*) immunohistochemically stained with a rat monoclonal antibody against the endothelial antigen PECAM. *c*, *In situ* hybridization of the control tumour with Flk-1 specific <sup>35</sup>S-labelled antisense probe; darkfield illumination. *d*, Brightfield illumination of an *in situ* hybridization with a *neo*<sup>r</sup>-specific antisense probe of the angiogenesis-inhibited tumour. The central necrosis of the tumour is surrounded by tissue with high expression of *neo*<sup>r</sup> and consisting of the virus-producing and infected cells.

**METHODS.** Immunohistochemistry and *in situ* hybridization on frozen sections has been described<sup>7</sup>. Signals were the same, but weaker, with antibodies against von Willebrand factor. For *in situ* hybridization, Flk-1 antisense cDNA comprising the 5'-2,619 bp of the receptor cDNA or an antisense cDNA of the *neo*<sup>r</sup> gene, respectively, were used as probes. Autoradiogram exposure was three days. There was no specific hybridization with the sense probes. N, necrosis. Bar represents 400  $\mu$ m (*a*-*c*) or 1 mm (*d*).



- Plate, K. H., Breier, G., Welch, H. A. & Risau, W. *Nature* **359**, 845-848 (1992).
- Shweiki, D., Itin, A., Soffer, D. & Keshet, E. *Nature* **359**, 843-848 (1992).
- Millauer, B. et al. *Cell* **72**, 835-846 (1993).
- Plate, K. H., Breier, G., Millauer, B., Ullrich, A. & Risau, W. *Cancer Res.* **53**, 5822-5827 (1993).
- Kashles, O., Yarden, Y., Fischer, R., Ullrich, A. & Schlessinger, J. *Molec. cell. Biol.* **11**, 1454-1463 (1991).
- Redemann, N., Holzmann, B., Wagner, E. F., Schlessinger, J. & Ullrich, A. *Molec. cell. Biol.* **12**, 491-498 (1992).
- Amaya, E., Musci, T. J. & Kirschner, M. W. *Cell* **66**, 257-270 (1991).
- Ueno, H., Colber, L. H., Escobedo, J. A. & Williams, L. T. *Science* **252**, 844-848 (1991).
- Miller, A. D. & Rosman, G. J. *Biotechniques* **7**, 980-988 (1989).
- Newman, P. J. et al. *Science* **247**, 1219-1222 (1990).
- Kim, K. J. et al. *Nature* **362**, 841-844 (1993).
- De Vries, C. et al. *Science* **255**, 989-991 (1992).
- Chen, C. & Okayama, H. *Molec. cell. Biol.* **7**, 2745-2752 (1987).
- Gorman, C. M., Gies, D., McCray, G. & Huang, M. *Virology* **171**, 377-385 (1989).
- Stewart, C. L., Schuetz, S., Vanek, S. & Wagner, E. *EMBO J.* **6**, 383-388 (1987).

**ACKNOWLEDGEMENTS.** We thank M. Longhi and L. Taylorson for assistance in tumour implantation experiments, E. Dejana for the anti-PECAM antibody, M. Clauss for <sup>125</sup>I-VEGF, and J. Arch for preparing the manuscript.



# Dead Romanovs identified by PCR

**Apart from the usual crop of gene assignments (for hereditary haemorrhagic telangiectasia, for example) this month's issue carries further the genetics of expanding repeating elements.**

THE use of DNA analysis for forensic purposes still resembles laboratory investigation in that those responsible cannot behave as automata. So much is clear from the genetic identification of the remains, found in a shallow grave 35 km west of Ekaterinburg two years ago, of Tsar Nicholas II, his wife Alexandra and three of their children, reported in the current issue of *Nature Genetics* by Peter Gill *et al.* (6, 130; 1994). Most of those concerned are from the British Government Forensic Science Service at Aldermaston, but Pavel L. Ivanov, from the Engelhardt Institute in Moscow, and Erika Hagelberg, from the department of biological anthropology at the University of Cambridge, have played important roles.

Whether it is called an execution or a murder, the killing of Nicholas II and his family is one of the haunting tales of the Russian Revolution. Historians have long since established that the parents, at least three of the five children, three servants and the family doctor were killed on 16 July 1918 by soldiers of the Bolshevik army. The plan that the bodies should be disposed of down a nearby mineshaft was frustrated by the breakdown of the truck in which they were being carried; instead, they were buried in a roadside grave. Russian forensic investigations of the skeletal remains have tentatively established that the grave did indeed contain the remains of the Tsar and much of his family. Their positive identification is an important, if macabre, extra.

How do you set out to identify skeletons that have been buried in the ground for three-quarters of a century? Best start with bone. One gram or thereabouts will yield 50 picograms of DNA, enough (if only just) for the purposes of PCR. Then do you analyse the chromosomal or the mitochondrial DNA? The former is less stable over time, the latter more plentiful and maternally inherited, and so more indicative of relatedness. In practice, you have no choice but to go for both. Gill *et al.* worked with five different short tandem repeats in the nucleotide sequence of the genome proper, and with two well-known hypervariable regions of human mitochondrial DNA. Sex determination of the bones has been done by amplification of the gene amelogenin that is common to the X and Y chromosomes.

Prince Philip, Duke of Edinburgh and husband of the present British Queen, is the great-grandson, by maternal lineage,

of the Tsarina's mother. Piquantly, a blood sample provided by the Duke is a crucial element in the identification of three of the skeletons at Ekaterinburg as those of sisters, and of Alexandra as their mother; all the mitochondrial sequences are identical. Ethicists will be intrigued that one of the first genetic sequences of a named and living individual to be published should be that of a member of a royal family.

The data also show that four of the nine skeletons in the grave at Ekaterinburg are not related to the Romanovs. They, presumably, are those of the doctor and the three servants. It follows that two of the children are missing from the grave. Gill *et al.* laconically note that the two concerned, the Tsarevitch, Alexei and his sister Anastasia, may have been burned or buried separately or have escaped.

The relationship between Nicholas II and his children is established by the similarity of the chromosomal DNA, which appears to show that the three children in the grave are the offspring of their parents, but Gill *et al.* stumble on an unexpected snag. The mitochondrial DNA from the Tsar's femur appears not to be chemically homogenous. Instead, for every mitochondrial DNA molecule with the base thymine (T) at position 16, 169 (in the standard notation), there are four molecules with cytosine (C) at the same place. This phenomenon, called "heteroplasmy", is a sign of recent mutation in one of possibly many mitochondrial DNA molecules. Plainly it is a matter of some importance for the forensic application of DNA analysis that such sources of confusion should be quickly recognized.

That, in a wider context, is one purpose of the article by Alec J. Jeffreys and five colleagues from the University of Leicester in the same issue (6, 136; 1994). They are concerned with the process of mutation in short tandem repeats, otherwise known as "minisatellites", in the human genome. One of these, known as MS32 (located on chromosome 1), consists of a stretch of DNA 29 base-pairs long that may be repeated between 12 and 800 times in different individuals, who will usually be endowed with minisatellites of different length by their two parents.

What determines the number of repeating units in a minisatellite? The question has an obvious bearing on the topical and even urgent question of the recently discovered role of repetitive trinucleotide

sequences in several heritable neurological diseases, where severity and/or the age of onset is determined by the number of repeating units. What Jeffreys *et al.* now find (by the analysis of sperm DNA) is that mutation seems to accompany meiosis, that addition of repetitive units is more common than subtraction, that the 3' end of a minisatellite is more liable to mutation than the other and that the mutation rate, while always relatively high, seems to vary between individuals.

What this implies for the origin of the genetic diseases based on mutations of the repetition number is not yet clear, but Robert I. Richards and Grant R. Sutherland have a model (6, 114; 1994). The idea is that the expansion of a triplet (or other repeat) occurs during DNA replication. If there is a break in the newly synthesized strand, the loose end may jump to mate with another complementary trinucleotide, whereupon DNA repair enzymes will lengthen the repetitive structure. For a structure that is already long, that chance that there will be two breaks within the repetitive element will be greater, whereupon catastrophic lengthening is more likely, which fits in with the way in which the observed mutation rate increases with length.

Expanded repetitive elements are also now known to occur in the somatic cells of certain tumours, colorectal cancer in particular. Stimulated by those observations and the knowledge that the expansion of a CAG repeat is associated with Kennedy's disease, R. Wooster *et al.* (6, 152; 1994) have surveyed a dozen known repetitive elements in the human genome, measuring the two alleles in normal and cancerous tissue.

The upshot of the survey is unspectacular but none the less interesting. Alleles expanded relative to those in normal tissue were found in only 16 of 196 patients. Only one case of an expanded dinucleotide repeat was found (in a breast cancer), expanded trinucleotide and tetranucleotide elements accounted for the others. In no case was more than one repetitive unit expanded and more than one allele (paternal or maternal) affected. What that implies is that the circumstances in colorectal cancer, (or at least the non-hereditary form of it) must be exceptional. And may the sprinkling of cases found by Wooster *et al.* imply that expanded elements are but symptoms, not causes, of the general run of cancers? ©